U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
2016

U.PORTO

DNA Methylation for Detection of Primary Colorectal Tumors

Micaela Faria Freitas

FC

DNA Methylation for Detection of Primary Colorectal Tumors

Micaela Faria Freitas

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Bioquímica

2016

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

DNA Methylation for Detection of Primary Colorectal Tumors

Micaela Faria Freitas
Master's Degree in Biochemistry
Department of Chemistry and Biochemistry
2016

Supervisor
Carmen de Lurdes Fonseca Jerónimo, Associate Professor, ICBAS, University of Porto

Co-Supervisor
Rui Manuel Ferreira Henrique, Guest Professor, ICBAS, University of Porto

IPOPORTO

U.PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

"*Nous avons trois moyens principaux : l'observation de la nature, la réflexion et l'expérience. L'observation recueille les faits ; la réflexion les combine ; l'expérience vérifie le résultat de la combinaison. Il faut que l'observation de la nature soit assidue, que la réflexion soit profonde, et que l'expérience soit exacte.*
On voit rarement ces moyens réunis"

Denis Diderot

Acknowledgements

I would like to acknowledge to my academic supervisors, Professor Carmen Jerónimo and Professor Rui Henrique, for their support and guidance throughout the duration of this project. For everything they taught me and for demanding more of me, making me a better scientist. To thank Professor Carmen for believe in me and give me a chance of doing my dissertation on her research group.

Very special thanks to Fernanda Silva from Department of Pathology for her major help in this project. I want to thank her for doing all the unstained slides and H&E stained slides needed in this project, and for all the good advises and support.

I want equally thanks to Doctor Sónia Carvalho from Department of Pathology for her help examining all the H&E stained slides and delimiting the tumor or normal area, depending on the case. For all the advice and other help provided. Thank you for your amazing support.

I want to thank Paula Lopes from Department of Pathology for her work doing the immunohistochemistry for all the CRC cases, which provides us with important information used in this project. I want to thank one more time to Professor Rui for examining all the immunohistochemistry slides in order to evaluate the protein expression.

I also want to thank Luís Antunes from Department of Epidemiology for his precious help in the statistical analysis needed in this project.

FCUP and ICBAS for providing the knowledge I needed for complete my master course.

Thanks to Cancer Biology & Epigenetics Group from Research Center of Portuguese Oncology Institute of Porto for providing the facilities and resources for the development of this project.

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (CI-IPOP-74-2016).

Aqui deixo um especial agradecimento a todos os meus colegas e amigos do Grupo de Biologia e Epigenética do Cancro que desde sempre me ajudaram a resolver as minhas dúvidas, contribuíram para discussões científicas fantásticas e construtivas

e acima de tudo pelo apoio. Quero agradecer em especial à Sofia e à Francisca pelos bons conselhos, pela ajuda em termos laboratoriais e por todo o conhecimento cientifico que partilharam comigo ao longo deste ano.

Às minhas 10 colegas de casa (Joana, Ana, Yara, Ana Rita, Carolina, Cristiana, Fabiana, Rita, Maria e Kati) que durante este ano foram a minha segunda família, por representarem um porto de abrigo, pelos momentos lúdicos que me proporcionaram, pela companhia e enorme apoio, nunca esquecerei.

Um enorme agradecimento à minha amiga Yara por ter sido minha confidente, por ter acreditado em mim quando nem eu acreditei e por me ter apoiado independentemente dos altos e baixos.

Às minhas amigas Joana e Rita quero agradecer por me terem aturado no desespero, por me terem feito acreditar que o impossível também é possível, por estarem comigo de forma incondicional e pelo apoio absoluto em todas as minhas decisões. Obrigada pela vossa amizade!

Por fim, quero agradecer à minha família: aos meus pais por acreditarem em mim e me apoiarem em todas as minhas decisões, por estarem sempre lá nos momentos difíceis, por me terem transmitido valores como humildade, responsabilidade e honestidade, e por me apoiarem de forma incondicional; ao meu irmão pelo seu apoio; ao meu avô António e avó Clara pelo imenso apoio, por acreditarem em mim, por serem avós babados, pelas discussões politicas com o meu avô, e à minha avó por me transmitir tranquilidade e me mostrar que a paz de alma existe; e finalmente não poderia deixar de agradecer ao meu gato Kiko pela companhia nas longas noites de estudo.

Obrigada a todos os que aqui foram mencionados e a todos os que não o foram, mas que de alguma forma também me apoiaram e acreditaram em mim, pois todos vocês contribuíram para o sucesso deste projeto.

Resumo

Introdução: Atualmente o cancro é a segunda doença que mais mata nos países ocidentais, sendo o cancro colo-rectal (CRC) um dos mais incidentes e mortais, excluindo o cancro da pele melanoma. De acordo com o modelo de progressão proposto por *Fearon* e *Vogelstein*, CRC resulta da acumulação de alterações genéticas e epigenéticas, resultando na transformação de epitélio glandular normal em adenoma, o que pode posteriormente propiciar a progressão para adenocarcinoma. As principais vias moleculares de carcinogénese associadas ao CRC são a instabilidade cromossómica (CIN), instabilidade microssatélite (MSI) e fenótipo de metilação de ilha CpG (CIMP). Ilhas CpG são regiões no genoma cuja percentagem de dinucleótidos CpG é superior ao que seria de esperar considerando uma distribuição aleatória de nucleótidos. Considerando o genoma humano, tem sido reportado que aproximadamente metade de todos os genes que codificam proteínas possuem uma ilha CpG na sua região promotora, a metilação nestas ilhas CpG pode induzir o silenciamento transcricional. Por esta razão, biomarcadores baseados na metilação de DNA podem ser utilizados tanto no diagnostico como no prognóstico de CRC. E apesar de já alguns testes baseados neste principio estarem a ser comercializados, a falta de consenso quanto à sua especificidade e/ou sensibilidade têm levantado dúvidas quanto á sua eficácia. Por este motivo, mais investigação é necessária no sentido de encontrar um bom teste de diagnóstico. A metilação em alguns genes tem sido reportada como um indicador global de cancro, como é o caso do *Adenomatous polyposis coli* (*APC*) e *Ras association domain family 1 - isoform A* (*RASSF1A*), presente em vários tipos de cancro, já a metilação em outros genes tem sido descrita como sendo bastante especifica para CRC, como são o caso do *MutL homolog 1* (*MLH1*), *O-6-methylguanine-DNA methyltransferase* (*MGMT*), *septin 9* (*SEPT9*) e *insulin-like growth factor 2* (*IGF2*). Assim um painel de biomarcadores constituído por alguns destes genes poderá apresentar uma boa especificidade e sensibilidade e desta forma ter potencial uso no diagnóstico e prognóstico do CRC.

Objetivo: Desenvolver um método não invasivo para detetar tumores primários de CRC baseado na avaliação nos níveis de metilação da região promotora de genes selecionados.

Métodos: Os níveis de metilação para os genes selecionados foi determinado para um total de 214 casos de tecido primário tumoral (CRC) e 50 casos de mucosa

normal (CRN). Todas as amostras foram obtidas a partir de tecido parafinado do IPO-Porto. Tanto as áreas tumorais como as de tecido normal, dependendo de se tratar de CRC ou CRN, foram dissecadas manualmente, seguindo-se a extração de DNA e tratamento bissulfito. Os níveis de metilação foram determinados com recurso a PCR quantitativo especifico de metilação (qMSP) usando SYBR Green I para a deteção do amplicon e *ACTβ* para a normalização.

Resultados: Os níveis de metilação no *SEPT9*, *RASSF1A*, *MGMT* e *APC* foram superiores no tecido tumoral, enquanto que no *IGF2*, foram menores, comparativamente com o observado para o tecido normal. Separando a análise pela localização do tecido, só o *SEPT9*, *MGMT* e *RASSF1A* para o colon e o *SEPT9*, *MGMT* e *APC* para o recto, apresentaram níveis de metilação significativamente superiores no tecido tumoral quando comparados com tecido normal. Para os tumores do recto, pacientes que fizeram tratamento neoadjuvante apresentavam níveis de metilação significativamente inferiores no *SEPT9* (*P*=0,002) e *MGMT* (*P*=0,012). Tanto para tumores no colon como no recto, foram observados níveis de metilação superiores em mulheres no *MGMT* (*P*=0,048 e *P*=0.049, respetivamente) e *MLH1* (*P*=0,007 e *P*=0,010, respetivamente) comparativamente com os níveis observados em homens. De forma geral, os genes selecionados foram capazes de detetar CRC em tumores independentemente da via molecular seguida, MSI ou CIN. Os genes com a melhor performance em termos de diagnostico foram o *SEPT9* com 85,5% de sensibilidade e 94,0% de especificidade, logo seguido pelo *MGMT*, com 77,2% de sensibilidade e 84,0% de especificidade. Já o painel com a melhor performance é constituído por *SEPT9/MGMT/RAASF1A* com 96.6% de sensibilidade, 74.0% de especificidade, PPV(valor preditivo positivo)=91.5% e NPV(valor preditivo negativo)=72.5%. A taxa positiva de deteção usada para avaliar a performance do painel selecionado entre as duas possíveis localizações do tumor e estadiamento, mostrou que o painel consegue detetar tumor em ambas as localizações e todos os estadios da doença com um valor sempre acima de 90%. Para além disso foi observado valor prognóstico para dois dos genes, sendo que a hipermetilação do *SEPT9* e *MLH1* mostraram estar significativamente associadas com um melhor prognóstico.

Conclusão: O painel com melhor acuidade diagnóstica foi o *SEPT9/MGMT/RASSF1A*, demostrando melhor sensibilidade, PPV e NPV que o *SEPT9* sozinho. O *MGMT* contribui com um enorme aumento na sensibilidade e NPV e o *RASSF1A* mantém esta tendência de aumento. Os resultados deste estudo suportam a

ideia de que o *MGMT* é um excelente biomarcador para ser utilizado em conjunto com o *SEPT9* em ensaios que visem o diagnóstico de CRC.

Palavras-chave: Cancro Colo-rectal; Metilação de DNA; Biomarcadores; Diagnostico; Prognostico;

Abstract

Background: Currently, cancer is the second most deadly disease in western countries and colorectal cancer (CRC) is one of the most incident and deadly, excluding non-melanoma skin cancer. According with commonly accepted model proposed by *Fearon* and *Vogelstein*, CRC results from the accumulation of both acquired genetic and epigenetic changes, which results in transformation of normal glandular epithelium into adenoma, that can progress to adenocarcinoma. According with molecular alterations, colorectal tumors are classified into three main pathways: chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylation phenotype (CIMP). CpG islands are regions in the genome where the percentage of the CpG dinucleotides is higher than would be expected, based upon a random distribution of nucleotides. It has been described that approximately half of all encoding-protein genes in the human genome contain a CpG island in their promoter region. As a consequence of their location in the promoter, methylation of CpG island may induce transcriptional silencing. Therefore, DNA methylation biomarkers might be used in diagnosis and prognosis of CRC. Although there are already DNA-methylation based tests being commercialized, their lack of specificity and/or sensitivity requires additional investigation. Previous reports have demonstrated potential usefulness of several genes, not only to detect cancer globally, such as *Adenomatous polyposis coli* (*APC*) and *Ras association domain family 1 - isoform A* (*RASSF1A*), and specifically to detect CRC such as *MutL homolog 1* (*MLH1*), *O-6-methylguanine-DNA methyltransferase* (*MGMT*), *septin 9* (*SEPT9*) and *insulin-like growth factor 2* (*IGF2*). Therefore, a panel of these genes might result in a robust DNA-methylation based test with good sensitivity and specificity leading to a potential use in diagnosis and prognosis of CRC.

Aims: To develop a non-invasive method to detect primary recurrent colorectal tumors based on the evaluation of promoter methylation of a selected panel of genes.

Methods: The methylation of CpG islands in the selected genes were analyzed in 214 cases of CRC and 50 normal controls (CRN). All the samples were obtained from paraffin-embedded tissues blocks from the IPO-Porto. The tumor and normal areas were dissected manually followed by DNA extraction and bisulfite conversion. The methylation status was determined by quantitative methylation specific PCR (qMSP) using SYBR Green I. The results were normalized using *ACTβ*.

Results: Methylation levels on *SEPT9*, *RASSF1A*, *MGMT* and *APC* were higher in tumor tissue, while in *IGF2*, higher methylation levels were found in normal tissue. When the tissue was separated by location, only *SEPT9*, *MGMT* and *RASSF1A* for colon, and *SEPT9*, *MGMT* and *APC* for rectum, showed significantly higher methylation for CRC tissues when compared to CRN. It was also observed, only for rectum, lower methylation levels in CRC tissues that went through neoadjuvant treatment for both *SEPT9* (*P*=0.002) and *MGMT* (*P*=0.012). Higher methylation levels were found in women, in both colon and rectum, for *MGMT* (*P*=0.048 and *P*=0.049, respectively) and *MLH1* (*P*=0.007 and *P*=0.010, respectively). In a generally way, selected genes were able to detect CRC in tumors either following MSI or CIN pathways. The genes with the best diagnosis performance were *SEPT9* with 85.5% sensitivity and 94.0% specificity, followed by *MGMT* with 77.2% sensitivity and 84.0% specificity. While the panel with the best performance was *SEPT9/MGMT/RASSF1A* with 96.6% sensitivity, 74.0% specificity, PPV=91.5% and NPV=72.5%. The positive detection rate used to evaluate the panel performance among the tumor location and stage, showed that the selected panel was able to detect tumors in both locations and all tumor stages (positive detection rates above 90%). Furthermore, some genes had prognosis value, hypermethylation in *SEPT9* and *MLH1* was associated with better prognosis.

Conclusion: The panel with the best diagnosis performance is *SEPT9/MGMT/RASSF1A* with better sensitivity, PPV and NPV than *SEPT9* alone. *MGMT* contributes with a huge increase in sensitivity and NPV, while *RASSF1A*, as a global cancer marker, contributes as well to the increase of panel sensitivity and NPV. Our data supports *MGMT* as a great biomarker to be used along with *SEPT9* in a DNA methylation assay for CRC detection.

Keywords: Colorectal cancer; Gene promoter methylation; Biomarkers; Diagnosis; Prognosis;

Contents

List of Acronyms and Abbreviations

A	***ACTβ***	*Actin β*
	AJCC	American Joint Committee on Cancer
	APC	*Adenomatous polyposis coli*
	AUC	Area under the curve
B	**BMI**	Body mass index
	BMPR1A	*Bone morphogenetic protein receptor type 1A*
	BRAF	B-Raf proto-oncogene, serine/threonine kinase
C	**C**	Cytosine
	CACNA1G	*Calcium voltage-gated channel subunit alpha1 G*
	CAP	College of American Pathologists
	CEA	Carcinoembryonic antigen
	CI	Confidence Interval
	CIMP	CpG island methylator phenotype
	CIN	Chromosomal instability
	CpG	Cytosine-phosphate-Guanine
	CRC	Colorectal Cancer
	CT	Computed Tomography
D	**DFS**	Disease-free survival
	DMR	Differentially methylated region
	DNA	Deoxyribonucleic acid
	DNMT	DNA methyltransferase
	dsDNA	Double-stranded DNA
	DSS	Disease-specific survival
E	**EGFR**	Epidermal Growth Factor Receptor
	ESMO	European Society for Medical Oncology
F	**F**	Forward
	FAP	Familial Adenomatous Polyposis
	FDA	Food and Drug Administration
	FFPE	Formalin-fixed paraffin-embedded
	5-FU	5-Fluorouracil
G	**G**	Guanine

	g	grams
	GTP	Guanosine triphosphate
H	**H&E**	Haematoxylin and eosin
	HNPCC	Hereditary Nonpolyposis Colorectal Cancer
	HR	Hazard Ratios
I	***IGF2***	*Insulin-like growth factor 2*
K	***KRAS***	KRAS proto-oncogene, GTPase
L	**lncRNA**	Long non-coding RNA
	LOI	Loss of imprinting
	LV	Leucovorin
M	**M**	Molar concentration (mol/L)
	MAC	Modified Astler-Coller
	MAP	MYH-associated Polyposis
	MAPK	Mitogen-activated protein kinase
	MGMT	*O-6-methylguanine-DNA methyltransferase*
	µL	microliter
	µm	micrometer
	MINT1	*Locus MINT1* used in CIMP panel
	MINT2	*Locus MINT2* used in CIMP panel
	MINT31	*Locus MINT31* used in CIMP panel
	miRNA	MicroRNA
	MLH1	*MutL homolog 1*
	MMR	Mismatch repair
	MRI	Magnetic Resonance Imaging
	MSH2	*MutS homolog 2*
	MSH6	*MutS homolog 6*
	MSI	Microsatellite instability
	MSI-H	MSI high
	MSI-L	MSI low
	MSP	Methylation specific PCR
	MSS	Microsatellite stable
	MYH	*MutY homolog*
N	**NCCN**	National Comprehensive Cancer Network

	NcRNA	Non-coding RNA
	NEUROG1	*Neurogenin 1*
	NOS	No other specification
	NPV	Negative predictive value
P	***p16***	*Cyclin dependent kinase inhibitor 2A (CDKN2A)*
	PCR	Polymerase chain reaction
	PET	Positron Emission Tomography
	PMS2	*PMS1 homolog 2, mismatch repair system component*
	pN	Pathologic N on TNM classification
	PPV	Positive predictive value
	pT	Pathologic T on TNM classification
	PTEN	Phosphatase and tensin homolog
Q	**qMSP**	Quantitative MSP
R	**R**	Reverse
	RASSF1A	*Ras association domain family 1 – isoform A*
	ROC	Receiver Operating Characteristic
	rpm	Rotations per minute
	RR	Relative Risk
	RUNX3	*Runt-related transcription factor 3*
S	***SEPT9***	*Septin 9*
	SMAD4	*SMAD family member 4*
	SOCS1	*Suppressor of cytokine signaling 1*
	STK11	*serine/threonine kinase 11*
	TP53	*Tumor protein p53*
T	***TS***	*Thymidylate synthase*
U	**UICC**	Union for International Cancer Control
V	**VEGF-2**	Vascular Endothelial Growth Factor-A
	VEGFR-2	Vascular Endothelial Growth Factor Receptor-2
W	**WHO**	Word Health Organization

Figure Index

Table Index

Introduction

Cancer

Normal cells evolve to a neoplastic state by acquiring a succession of biological properties, also known by “hallmarks of cancer” (Figure 1). At the present, a total of ten hallmarks are acknowledged, including: sustaining proliferative signaling, cell death-resistance, growth suppressors evasion, immune destruction avoidance, angiogenesis, replicative immortality, tumor promoting inflammation, genome instability and mutation, deregulated cellular energetics and invasion and metastatic capacities [1].

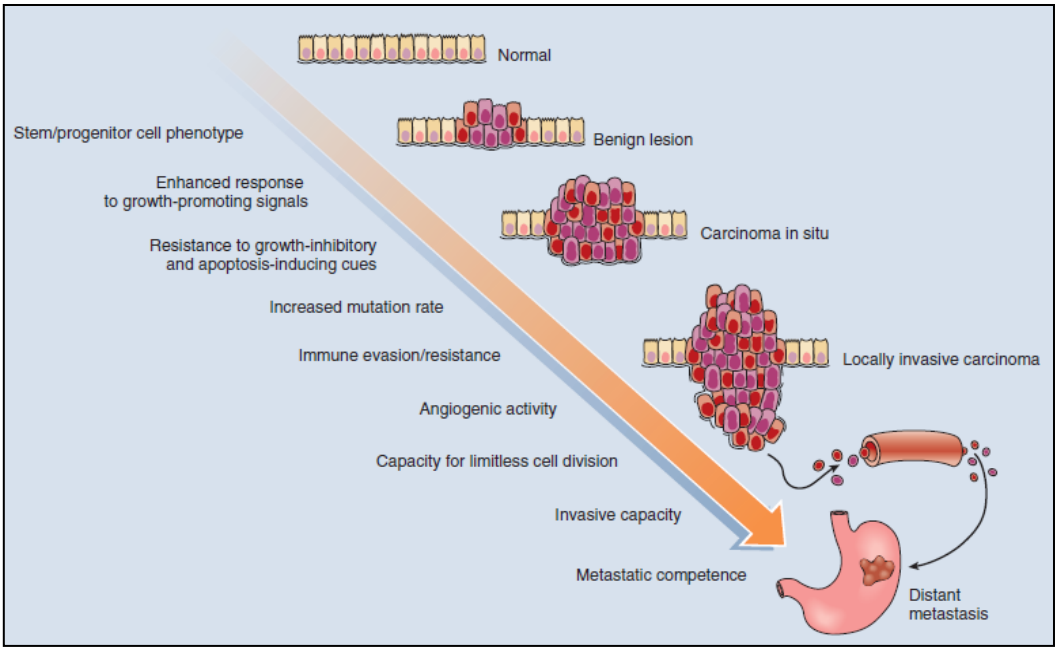

Figure 1 – Scheme pointing out some of key traits involve in cancer progression. This represents a typical epithelial cancer, such as those that typically arise in the lung, colon, breast, or prostate [2].

Colorectal cancer: Incidence, risk factors and treatment

Incidence

Cancer is, currently, the second most deadly disease [3]. Colorectal cancer (CRC) is one of the most incident and deadly type of cancer, excluding non-melanoma skin cancer (Figure 2). CRC is the third most incident and the fourth most deadly type of cancer worldwide [3]. Being the third most common cancer in men and second in women.

Regarding mortality, it is ranked in third position for females and in the fourth in males [4].

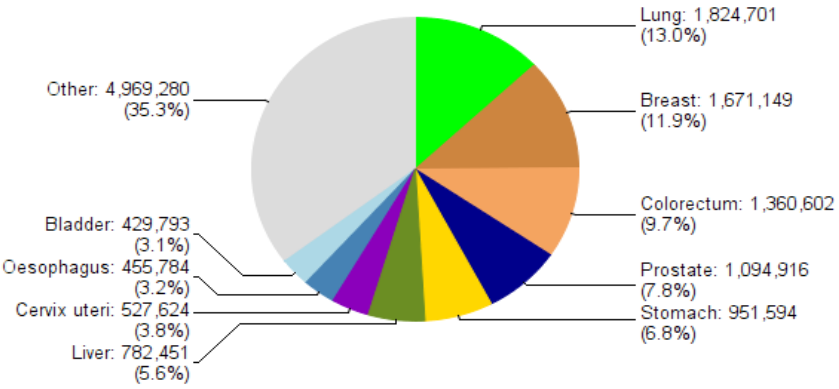

Figure 2 – World's incidence of different types of cancer for both sexes, all ages. Estimated number of cancer cases in 2012. (Adapted from [4])

In Europe, in 2012, there were estimated 3,45 million new cases of cancer (excluding non-melanoma skin cancer) and 1,75 million deaths due to cancer [5]. CRC was the second most incident and deadly type of cancer. Whereas, for woman, CRC was the second most incident and deadly, for man was the third most incident and second most deadly [5] (Figure 3).

Considering our country, both in man and woman, CRC is the second most incident and deadly type of cancer following the European trend [5] (Figure 3).

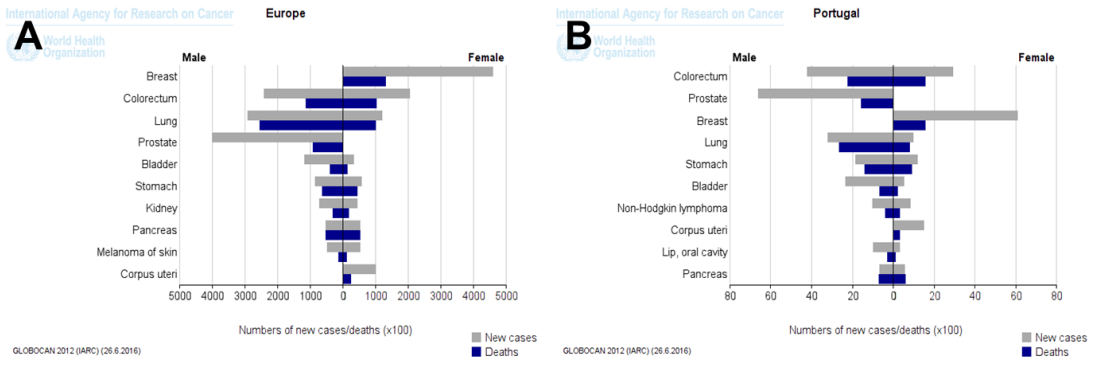

Figure 3 – Incidence and mortality for the ten most incident types of cancer in Europe (A) and Portugal (B) by gender [5].

Types of CRC

The majority of colorectal cancers are adenocarcinomas. But this is not the only primary type of cancer found in this location. Besides adenocarcinoma, other types of cancer are found, specifically, the gastrointestinal stromal tumor, the carcinoid tumor, the

squamous cell cancer of the anus, and the lymphoma. The first two types are, respectively, developed from interstitial cells of Cajal and gastrointestinal neuroendocrine cells. Regarding the squamous cell cancer of the anus, it is associated with infection by human papilloma virus. [6].

Considering the histologic types of CRC, they are categorized according to the classification proposed by the World Health Organization (WHO), recommended by the College of American pathologists (CAP). The types included in this classification are: adenocarcinoma; carcinoma no other specification (NOS); Adenocarcinoma *in situ*; cribriform comedo-type, micropapillary and serrated adenocarcinomas; medullary, mucinous, signet ring cell, squamous cell (epidermoid), spindle cell, and adenosquamous carcinoma; neuroendocrine carcinoma (NEC), large NEC and small NEC; undifferentiated carcinoma; and mixed adenoneuroendocrine carcinoma [7, 8]. Some examples are displayed on Figure 4 considering normal colon (A) as reference.

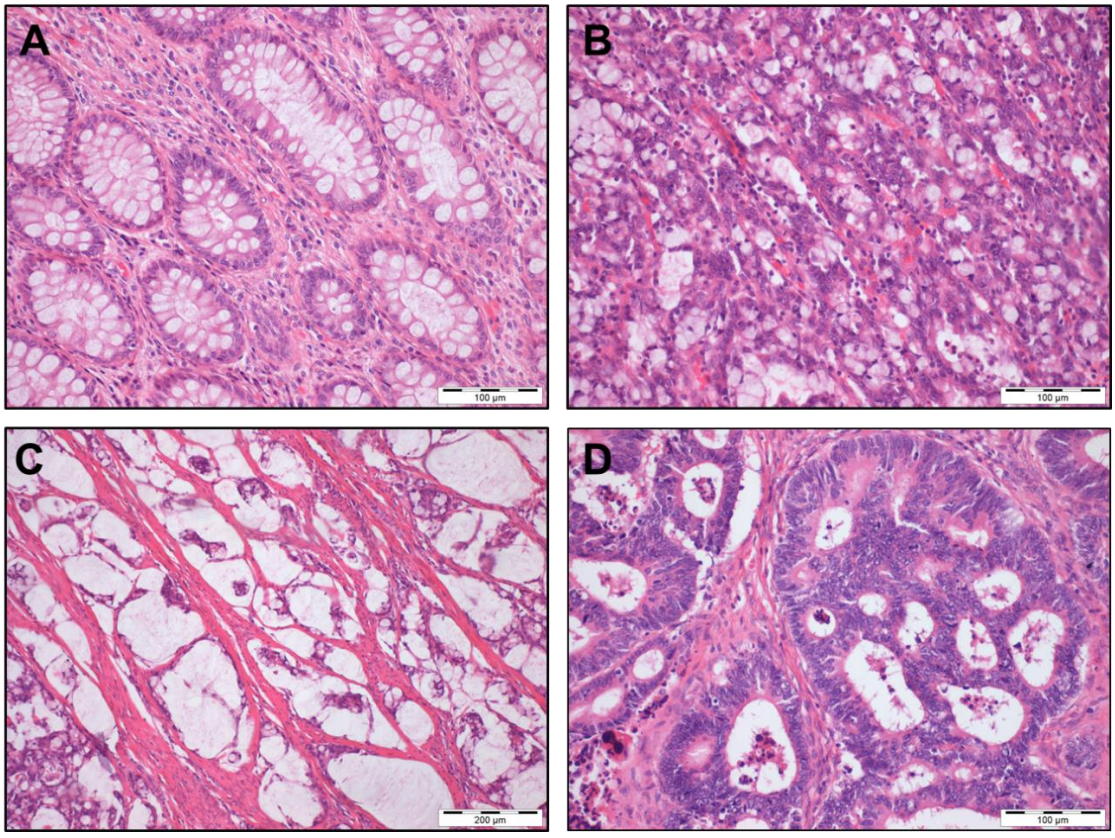

Figure 4 – Mucosa histology. A, Normal colon (from patient CRC176). B, Adenocarcinoma with neoplastic cells displaying mucus production (from patient CRC267). C, Mucus with neoplastic cells (from patient CRC267). D, Adenocarcinoma displaying a cribriform architecture (from patient CRC383).

At a superficial level CRC can be classified according to its origin into sporadic or hereditary. Sporadic CRC is a somatic genetic disease without known contribution

from germline causes, significant family history of cancer or inflammatory bowel disease. In other hand, hereditary CRC is due to germline alterations [9].

Risk Factors

The risk factors for CRC can be divided into two categories: genetic and environmental. The genetic risks are due to personal and family history in which genetic characteristics are implicated. While the environmental risks are related with the diet, sedentary lifestyle, obesity, smoking and alcohol [6, 8]. All these risk factors might interact with each other, leading to initiation and development of the disease [10].

Considering the family history, first-degree relatives with CRC and inherited syndromes are the main appointed risk factors. There are several syndromes associated with CRC, the two most common are hereditary nonpolyposis colorectal cancer (HNPCC), also known by Lynch syndrome or cancer family syndrome, and familial adenomatous polyposis (FAP). The other syndromes caused by germline mutations are MYH-associated polyposis (MAP), Cowden syndrome, Juvenile polyposis, Peutz-Jeghers syndrome and Bannayan-Ruvalcaba-Riley syndrome [10, 11]. However, the genetically inherited syndromes of CRC are very rare (3-5%). Moreover, high incidence of CRC can be observed in a given family without a genetically inherited syndrome been involved [6]. Persons who have adenoma in first-degree relatives have an increased risk of colorectal cancer, with a relative risk (RR) of 4.38 (95% confidence interval (CI), 2.25-8.43) when compare to persons who don't have this personal history [12].

Regarding the personal history, the risks are age higher than 40 years old, a history of previous CRC or inflammatory bowel disease [13], male gender and diabetes mellitus [8, 11]. Indeed, it has been described that men have a higher risk of developing both colon (OR=1.4) and rectal cancer (OR=1.7) when compared with women [6, 14]. Diabetes was associated with an increased risk of CRC, compared with no diabetes, RR=1.26 (95% CI, 1.20-1.31), considering a meta-analysis of 24 studies [15].

Risk factors associated with the diet are still controversial. It has been generally accepted that a diet high in fat, red meat and processed meat increases the chances of developing CRC [16]. Considering a dose-response meta-analyses, red and processed meat intake (for 100g/day increase) was significantly related to an increased risk of colorectal, RR=1.14 (95% CI, 1.04-1.24), colon with a RR of 1.25 (95% CI, 1.10-1.43), and rectal cancer with a RR=1.31 (95% CI, 1.13-1.52) [17]. However, the effect of fiber intake in CRC has been extensively debated. Some studies showed that the high intake of fiber have a protective effect on CRC, while other studies fail to reach the same conclusions [6].

As mentioned earlier, smoking increases the risk of CRC, considering twenty-six studies pooled together, in a meta-analysis, smoking has a RR of 1.18 (95% CI, 1.11-1.25) comparing ever smokers to never smokers [18].

The alcohol intake increases the risk of CRC, comparing with non-drinkers, the pooled multivariable RR were 1.16 (95% CI, 0.99-1.36) for persons who consumed between 30-45 g per day and 1.41 (95% CI, 1.16-1.72) for those who consumed more than 45 g per day [19].

Considering the increased risk of CRC with obesity, a 5 unit increase in body mass index (BMI) was related with an increased colon cancer risk in man, RR=1.30 (95% CI, 1.25-1.35) and woman, RR=1.12 (95% CI, 1.07-1.18). Despite this, in rectal cancer, BMI was associated with an increased risk in man, RR=1.12 (95% CI, 1.09-1.18), but this association was not found in woman, RR=1.03 (95% CI, 0.99-1.08) [20]. Herein, the relation between obesity and colon and rectal cancer risk varies by gender and cancer site [20].

Diagnosis

Several aspects have to be taken into account for diagnosis. More advanced stages of CRC have a few symptoms associated while a yearly-stage for not having those can go unnoticed for a while. The main symptoms associated with CRC are rectal bleeding, weight loss, abdominal pain, diarrhea, constipation, abnormal rectal examination, abdominal tenderness, haemoglobin <10.0 g/dL, positive faecal occult blood and blood glucose >10 mmol/L [21].

Screening is made in the general population in order to detect a pre-cancer condition or a very early-stage malignancy before it being symptomatic [22]. The most frequent examinations used are colonoscopy, guaiac-based faecal occult blood test, imaging to help identify metastatic disease, where it can be included computed tomography (CT), magnetic resonance imaging (MRI), x-ray and positron emission tomography (PET), and finally, endorectal ultrasound [23]. For a definitive diagnosis, a tumor biopsy is performed in order to obtain a histological confirmation.

Staging

After the CRC diagnosis, clinical examination, laboratory tests and instrumental search for metastasis should be performed in order allow a complete staging.

The disease stage is described according to the TNM system, following the guidelines of American Joint Committee on Cancer (AJCC)/ Union for International

Cancer Control (UICC) as shown in Table 1. In TNM, T stands for Primary Tumor, N for Regional Lymph Nodes, and M for Distant Metastasis. To evaluate the microscopic invasion (T), it is considered the histologic layers displayed on Figure 5. Further, TNM stage can be written with the prefix letters c, p, y or r, where the c stands for a clinical classification, p for the pathologic one, y for those cancers that were classified after neoadjuvant pretreatment and the r is used for those cancers that had recurred after a disease-free interval [7].

Table 1 – TNM classification of colorectal cancer according to AJCC/UICC guidelines (Adapted from [7]).

T - Primary Tumor	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ: intraepithelial or invasion of lamina propria
T1	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades through the muscularis propria into pericolorectal tissues
T4a	Tumor penetrates to the surface of the visceral peritoneum
T4b	Tumor directly invades or is adherent to other organs or structures
N – Regional Lymph Nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1-3 regional lymph nodes
N1a	Metastasis in one regional lymph nodes
N1b	Metastasis in 2-3 regional lymph nodes
N1c	Tumor deposit(s) in the serosa, mesentery, or nonperitoneal tissues without regional nodal metastasis
N2	Metastasis in four or more regional lymph nodes
N2a	Metastasis in 4-6 regional lymph nodes
N2b	Metastasis in seven or more regional lymph nodes
M – Distant Metastasis	
M0	No distant metastasis
M1	Distant metastasis
M1a	Metastasis confined to one organ or site
M1b	Metastases in more than one organ/site or the peritoneum

Each case can be staged for several times, normally the first time is at diagnosis. The first staging is very important since it will define the first line treatment. Then, it can be stage after any treatment, chemotherapy/radiotherapy or even surgery. Hence, it is possible to assess the disease progression, allowing the necessary treatment adjustments upon the new TNM classification [7].

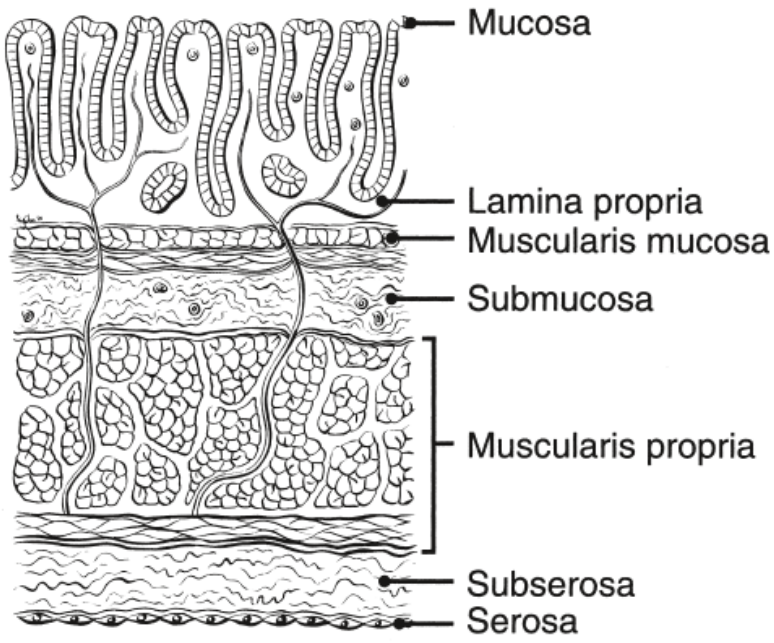

Figure 5 – Tissue layers of colon (Adapted from [7]).

The anatomic staging and histology grade should be determined as well (Table 2). Histologic grade can be classified into low grade (Well/Moderately differentiated) and high grade (poorly differentiated and undifferentiated).

Table 2 – Anatomic Stage/Prognostic Grouping with the correspondent TNM classification (at the left side). And Histologic grade (at the right side) (Adapted from [11])

Stage	TNM Correspondence	Histologic Grade	
I	T1-2, N0, M0	GX	Grade cannot be assessed
IIA	T3, N0, M0	G1	Well differentiated
IIB	T4a, N0, M0	G2	Moderately differentiated
IIC	T4b, N0, M0	G3	Poorly differentiated
IIIA	T1-2, N1, M0	G4	Undifferentiated
IIIB	T3-4, N1, M0		
IIIC	T1-4, N2, M0		
IVA	T1-4, N1-2, M1a		
IVB	T1-4, N1-2, M1b		

In order to have a complete staging, the tumor location is needed. In Figure 6 the anatomic division of the colon, from the appendix until the anus is displayed.

The AJCC/UICC classification system is most commonly considered international standard for CRC staging system. However, others like Dukes and the Modified Astler-Coller classification (MAC) can be alternatively used for staging [8].

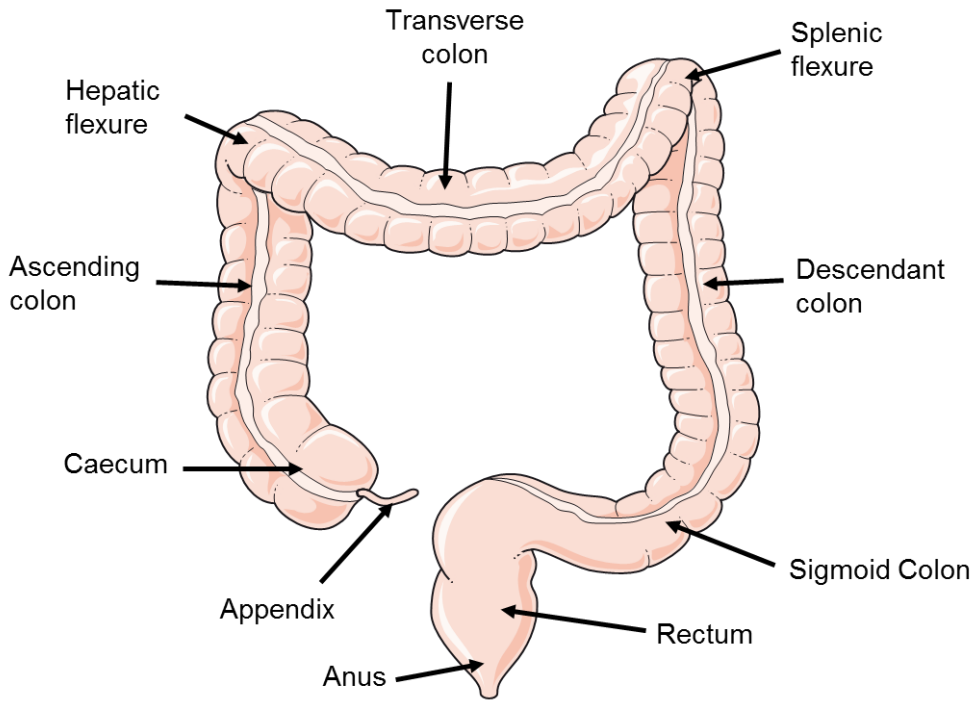

Figure 6 – Anatomic subsites from appendix to anus.

After primary surgical resection, the postoperative residual disease status of the patient should be evaluated. This status reflects the efficacy of primary treatment and is categorized by the R classification system (Table 3). This classification has a stage-independent prognostic value, where, at any stage of disease, R1 and R2 are adverse prognostic factors when compared to R0. The R0 code refers not only to the absence of primary tumor but also to the absence of tumor dissemination, namely the absence of distant metastasis [7, 8].

Table 3 – R classification system. (Adapted from [7])

RX	Presence of residual tumor cannot be assessed
R0	Complete tumor resection with all margins histologically negative
R1	Incomplete tumor resection with microscopic residual tumor
R2	Incomplete tumor resection with macroscopic residual tumor

Besides the postoperative residual disease status, when the patient is submitted to preoperative adjuvant (neoadjuvant) treatment, the tumor regression grade should also be appraised. This evaluation is made according to the CAP guidelines, consisting in a four-point tumor regression grade as shown in Table 4 [7].

Table 4 – Tumor regression grade (Adapted from [7]).

Tumor regression grade	Description
0 (Complete response)	No viable cancer cells
1 (Moderate response)	Single cells or small groups of cancer cells
2 (Minimal response)	Residual cancer outgrown by fibrosis
3 (Poor response)	Minimal or no tumor kill; extensive residual cancer

Treatments

An adequate staging will guide the adequate treatment. Depending on the tumor location, rectum or colon, and time of detection/diagnosis, in addition to other factors, the therapy approach will be different. The treatment selection is a very complex process that will depends on multiple factors. Current treatment guidelines, are proposed by European Society for Medical Oncology (ESMO) and National Comprehensive Cancer Network (NCCN), which time to time are reviewed in the light of new knowledge [22, 24-28].

According to ESMO, the treatment should be risk-adapted. Briefly, in rectal cancer, for the earliest and most favorable cases, a surgery involving the tumor resection should be enough, if the margins are tumor free (R0). When surgery is contraindicated, a treatment based on chemo-radiotherapy, in which local radiotherapy used alone or combined with neoadjuvant chemotherapy is more suitable. For intermediate cases preoperative radiotherapy, followed by local tumor resection is suggested. Lastly, for the locally advanced tumors, which are mostly not resectable, the first treatment is chemo-radiotherapy followed by radial surgery. In some specific situations, like stage III patients (and “high-risk” stage II) it is given adjuvant chemotherapy as a postoperative treatment [26].

In colon cancer, stage I patients are treated with surgical resection and anastomosis, normally no adjuvant chemotherapy is needed. For stage II patients, the first treatment recommended is surgical resection and anastomosis, followed by adjuvant therapy for high-risk patients. Stage III patients besides the surgical resection and anastomosis, are treated with adjuvant chemotherapy [22].

Both in metastatic colon and rectal cancer, the first-line of treatment is systemic. The chemotherapy will have a curative intention if the metastasis is potentially resectable, otherwise, it will be palliative [28].

There are several different regimens of chemotherapy and subsequent different combinations. The regimens could be composed by only one drug (monotherapy) or

combined. The drugs normally used are leucovorin (LV) (also known as folinic acid, Isovorin or Levofolinate), 5-fluorouracil (5-FU), Oxaliplatin (Eloxatin), Irinotecan (Camptosar), and Capecitabine. The most commonly used combined regimens are FOLFOX, FOLFORI, XELOX (also known by CAPOX), XELIRI and FLOX (also known by *de Gramont*) (Table 5). Additionally, other combinations may exist and there are several new drugs being tested in clinical trials [24, 25].

Table 5 – Most common regimens used in chemotherapy, name and composition.

Regimens	Composition
FOLFOX	5-FU, LV and Oxaliplatin
FOLFORI	5-FU, LV and Irinotecan
XELOX	Oxaliplatin and Capecitabine
XELIRI	Irinotecan and Capecitabine
FLOX	5-FU and LV

Moreover, chemotherapy may be combined with biologic agents. Biologic agents are drugs with a very local specific action, normally a component of a signaling pathway. The most well-known biologic agents currently been used are Bevacizumab, Cetuximab and Panitumumab. Bevacizumab is an antiangiogenic monoclonal antibody directed against all isoforms of vascular endothelial growth factor-A (VEGF-A), thus disrupting their interaction with vascular endothelial growth factor receptor-2 (VEGFR-2). Cetuximab and Panitumumab are very similar monoclonal antibodies that specifically target the extracellular domain of Epidermal growth factor receptor (EGFR), preventing a ligand-induced activation, promoting receptor internalization and degradation [8].

Follow-up

During cancer treatment or after completing, follow-up is a standard practice. It consists in periodic consultations in order to prevent recurrence and new cancer, assess medical and psychological late effects, and lastly promote a healthy lifestyle [22]. The tests used for patient's follow-up are mostly, the same ones mentioned for diagnosis. In addition, serum marker carcinoembryonic antigen (CEA) testing might be used. The CEA is also assessed for response evaluation to adjuvant chemotherapy treatment [28]. Briefly, the tests commonly used in follow-up are CEA testing, full colonoscopy and CT scan of the thorax with abdomen and pelvis. Other tests can be performed in order to obtain supplementary information. The frequency of follow-up exams will depend on the risk associated to each patient and on the type of test that will be performed [8].

Genetic and epigenetic of colorectal cancer

Genetic and CRC

The genetic alterations in CRC are well characterized, they involve changes in the deoxyribonucleic acid (DNA) sequence altering the gene product by changing the amino acid sequence of protein or by altering the quantity of protein produced [29]. Depending on the original gene function, the outcome will be different. In some specific *tumor suppressor genes*, the defects can involve loss-of-function while in certain *oncogenes*, it will involve the gain-of-function [9]. Besides the alterations observed in *oncogenes* and *tumor suppressor genes*, alterations in genes regulating transcription and translation, as well as in genes responsible for DNA repair, may occur as well [2]. The genetic alterations include copy-number alterations, point mutations, deletions, and translocations, being most of them immaterial to neoplasia [30]. Nevertheless, some of these somatic defects might confer a selective growth advantage to the pre-neoplastic cell and this way, they would be “driver” events leading to malignancy. It has been identified, in the human genome, about 138 driver genes, being 74 tumor suppressor genes and 64 oncogenes. Although, a sporadic CRC might only contain 2-8 driver gene alterations [9, 30].

As mentioned before, a small portion (3-5%) of CRC are hereditary, on Table 6 are described the target genes mutated in some of those inherited disorders [10].

In addition to mutations associated with inherited conditions, there are others that have been described to play an important role as well. These mutations, beyond being relevant in terms of prognosis, are also, very informative in terms of therapy response [23]. For this reason, some of them are currently used in clinic as predictive markers. The result of this mutational analysis will define the best treatment scheme. Specifically, mutations at codons 12 and 13 of *KRAS* and at V600E (Valine 600 to glutamic acid) of *BRAF* are indicative of resistance to anti-EGFR monoclonal antibody therapy. Namely, KRAS mutations reduce GTPase activity, leading to constitutive activation of downstream pathway, whereas V600E leads to constitutive BRAF activation. Alterations in the enhancer region of *Thymidylate synthase* (*TS*), may also alter TS expression, thus, reducing or increasing patients’ response to fluorouracil [8, 23].

Table 6 – Some of the inherited disorders, and its variants, causing CRC [10].

Inherited Condition	Genes Mutated
HNPCC (Attenuated HNPCC, Muir-Torre syndrome, Trimbath syndrome, Turcot syndrome)	Mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6* and *PMS2*.
FAP (Attenuated FAP, Crail syndrome)	*APC*
MAP (MYH-associated colon cancer)	*MYH*
Juvenile polyposis	*SMAD4*, *BMPR1A*
Cowden syndrome	*PTEN*
Bannayan-Ruvalcaba-Riley syndrome	*PTEN*
Peutz-Jeghers syndrome	*STK11*

Epigenetic and CRC

Epigenetics involve gene expression regulation without changes in the DNA sequence and, consequently, are potentially reversible [29]. The understanding of human epigenome is much less detailed than the genome. This might be due to the fact that there are almost as many epigenomes as cell types, despite the slight differences between tissues, the fraction of changes have a huge impact in both differentiation and disease [31]. The epigenetic mechanisms include DNA methylation, histone post-translational modifications, non-coding RNAs (NcRNA), and histone variants (Figure 7). CH Waddington was the first to introduce the concept of epigenetics in 1939, but only in 1983 it was associated with cancer [31].

CRC is one of the types of cancer in which, over the years, consistent patterns of epigenetic alterations, has been described. The most frequently described alteration is only present in a subgroup of CRC, and it consists in the promoter hypermethylation of several genes leading to their inactivation. This alteration is also known by CpG (Cytosine phosphate Guanine) island methylator phenotype (CIMP) [2]. Besides CIMP, other epigenetic alterations can be found in CRC, such as, aberrant histone modification changes, and microRNA (miRNA) and long noncoding RNAs (lncRNA) deregulated expression [32].

CpG islands are genomic regions in which the percentage of the CpG dinucleotides is higher than it would be expected (based upon a random distribution of nucleotides). These regions are defined as 500-1000 bases in length with greater than 50% CG (cytosine and guanine) content and ratio of observed to expected CpGs of >0,6 [33]. Cytosine methylation is catalyzed by DNA methyltransferases (DNMTs). The methylation maintenance during replication is ensured by DNMT1, while *de novo*

methylation is catalyzed by DNMT3A and DNMT3B [34]. The DNA is regulated not only by DNMT but also by transcription factors. Several factors may contribute to aberrant DNA methylation. These are divided into intrinsic and extrinsic and whereas intrinsic factors include aging and mutations, extrinsic factors include diet, intake of mutagenic chemicals and smoking [34].

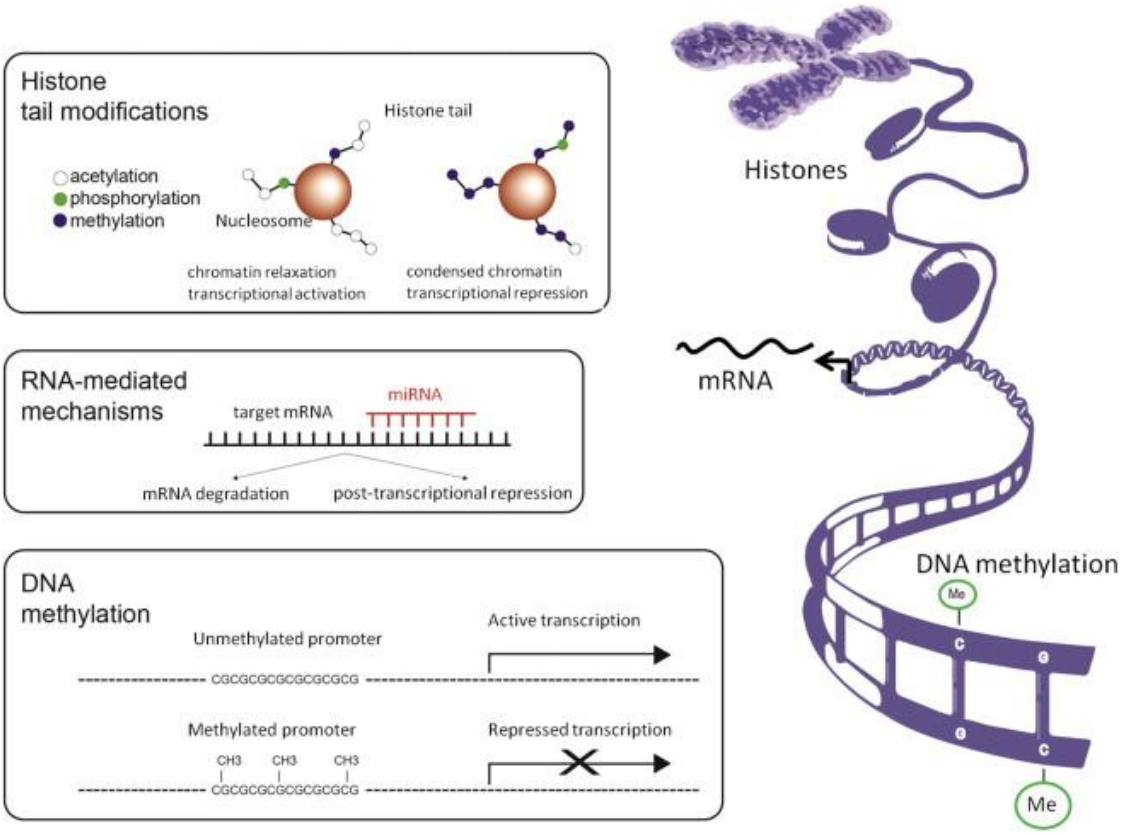

Figure 7 – Scheme with the main epigenetic mechanisms that can be deregulated in cancer [35].

There are two kinds of methylation in the human genome, the CpG and the non-CpG methylation. The non-CpG methylation is mostly described to be present in embryonic stem cells (25%) and its disappearance has been reported to occur upon differentiation inducement. While CpG methylation is present in all kinds of cells, stem and differentiated [36].

It is well known that approximately half of all encoding-protein genes in the human genome contain a CpG island in their promoter region [32, 37]. As consequence of their location in the promoter, CpG islands methylation associate with transcriptional silencing [32, 38]. Therefore, the biological outcome will differ depending on where the DNA methylation took place. In the other side, the absence of DNA methylation does not imply promoter activity or increase of gene transcription. It only means that the gene becomes regulated by other processes [31, 39].

In normal cells, the CpG island promoters are normally methylation-free, only a few (3%) are methylated. Besides CpG island promoters, there are non-CpG island promoters, or CpG-poor promoters, that in normal tissues are found typically methylated. The relation between methylation and expression is not so clear in the last ones [31]. Conversely, cancer cells have a global genome hypomethylation and a gene promoter-associated hypermethylation [39].

Over the years several DNA methylation targets has been reported, non-CIMP related, both at promoter and intronic regions, that were able to distinguish between healthy and malignant tissues. Although not all of them were able to correctly predict CRC, and so, only a few of these markers were approved and advanced to clinical trials [34]. This comes from the fact that *de novo* methylation of promoter CpG islands is a very frequent alteration observed in cancer [31].

There are epigenetic driver alterations that lead to malignancy as well as described for genetic alterations, one example is the hypermethylation of *MutL homolog 1* (*MLH1*). The promoter methylation of this gene leads to its inactivation and is present in almost all sporadic CRC with microsatellite instability (MSI) [31].

Molecular and morphological pathways in CRC

CRC results from the accumulation of both acquired genetic and epigenetic changes, which results in transformation of normal glandular epithelium into invasive adenocarcinoma.

As mentioned earlier, colorectal tumors might be classified according with various clinical, pathological and/or molecular features. Recently, molecular classification has been gaining importance because it reflects underlying mechanisms of carcinogenesis [40].

The main carcinogenesis pathways implicated in CRC involve chromosomal instability (CIN), MSI and CIMP. Notwithstanding the importance of molecular pathways, the underlying the tumor morphology are also important and should not be disregarded. Indeed, there are two main morphology pathways to be considered, the adenoma-carcinoma pathway and the serrated neoplasia pathway. These pathways might include alteration from all of the three molecular pathways.

Molecular pathways: CIN, MSI and CIMP

There are two levels of genetic instability, a subtle one affecting only DNA sequences, MSI-high (MSI-H), and a grossly one, affecting portions or entire chromosomes, the CIN. These forms of instability are reported to be mutually exclusive, thus a CRC with CIN is most likely microsatellite stable (MSS) [41, 42].

CIN pathway was the first one to be described, although the underlying mechanism is still not known. This is the most common molecular pathway, affecting proximally 85% of sporadic CRCs. Cohorts of CRCs with CIN display several karyotype aberrations, which includes chromosome gains and losses, chromothripsis, chromosome rearrangements, focal gene amplifications and at even base substitutions and deletions [9]. The neoplastic transformation normally occurs through this pathway by copy number gains or losses of driver genes [40]. Several mechanisms potentially responsible for CIN have already been proposed, however neither of them were consistently observed [9].

MSI refers to altered lengths (“instability”) of short nucleotide repeat sequences (“microsatellites”) in tumorous DNA when compared to normal DNA. This pathway is considered an alternative one to CIN. MSI-H, observed in approximately 15% of sporadic CRCs, has been reported to be due to defects in the DNA mismatch repair (MMR) system [9, 40]. DNA MMR system is responsible by the in recognition and direct repair of nucleotide mispairs right after replication, thus ensuring two identical daughter cells originated by mitosis. When the system fails, DNA fidelity is lost and mutations are transmitted in an asymmetric manner to both daughter cells and consequently multiple mutations will be, eventually, accumulated in the newly formed cells. In order to assess the MSI status, a panel of microsatellite markers, defined by National Cancer Institute, need to be tested. MSI-H tumors harbour more than 30% of frameshift mutation in microsatellite markers [9].

In sporadic CRC, the most common cause for defects in DNA MMR is aberrant bi-allelic hypermethylation of *MLH1*, a DNA MMR gene. *MLH1* is one of the markers used to define CIMP, thus, there is an overlap between many CIMP tumors and MSI-H tumors [9]. Conversely, MSI-L has been associated with *O-6-methylguanine-DNA methyltransferase* (*MGMT*) methylation and its correspondent loss of expression [40, 43].

As earlier mentioned, CIMP occurs in 10-40% of the patients resulting from the concomitant hypermethylation of selected loci depending on CIMP definition used [44]. CIMP is of high importance, once it affects both tumor suppressor and DNA repair genes.

Despite the intensive research in this field, there are no consensus guidelines on definition of CIMP. Moreover, there is no CIMP-gene panel robust enough to be generally accepted. Similarly, there is no unanimity in the marker thresholds or techniques used for detecting altered DNA methylation used to define this phenotype. This lack of consensus hampers the conclusion whether the difference in CIMP prevalence between studies is due to different methodologies, or in the primers' choice and/or location of methylation in the markers [45]. Importantly, the biological cause of CIMP remains unknown.

The categories of CIMP are an additional topic of debate. Some authors claim that it should be distinguished as two categories, "CIMP" and "non-CIMP" [46], others in three categories, either "CIMP-high, CIMP low, CIMP-0" [47] or "CIMP1, CIMP2, CIMP-negative" [48], and most recently, four categories, CIMP-high, CIMP-low and two clusters of non-CIMP depending on the frequency of *TP53* mutation [38].

Nevertheless, it has been shown that CIMP-high is associated with proximal tumor location, female sex, poor differentiation, MSI, high *BRAF* and low *TP53* mutation rates [44, 49-52]. Conversely, CIMP-low is associated with KRAS mutation and male sex [47], and CIMP-0 (CIMP-negative) is associated with wild-type *BRAF/KRAS* and a distal tumor location [40, 53].

Currently there are two different CIMP panels being used for CIMP status assessment, the classical and the novel. The classical panel includes *MLH1*, *MINT1*, *MINT2*, *MINT31*, and *p16* [54], while the novel panel is composed by *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS1* [46].

Morphological pathways

There are several forms of colorectal epithelial polyps, the adenoma and the serrated polyps which include hyperplastic polyps, traditional serrated adenomas and sessile serrated adenomas [55]. Adenomas were the first to be perceived to represent precursor lesions to CRC, with the first description of the classic adenoma-carcinoma progression model proposed by Fearon and Vogelstein [56]. Since then a lot of effort was made in the understanding of the biology of this malignancy resulting in successive revisions of the initial model. Hyperplastic polyps, in the other hand, were considered to have no potential for progression to malignancy, but recently was recognized that some polyps have a significant risk for neoplastic transformation. Thus, progressing to CRC through another pathway: the serrated neoplasia pathway [55].

The adenoma-carcinoma pathway stands that the stepwise change in morphology and accumulation of molecular alterations is a concomitant event.

Considering the morphology, it starts with the formation of a small adenomatous polyp, followed by progression to a larger polyp with dysplasia, which eventually will progress to invasive carcinoma [55, 56]. In this pathway, the first molecular abnormality to arise is the mutation of *Adenomatous polyposis coli* (*APC*), which is followed by the acquisition of mutations in *KRAS* and *TP53* [55, 57]. This pathway responsible for sporadic CRC is also shared by FAP [55].

The serrated pathway is responsible by 30% of CRC and has as end-point the serrated adenocarcinoma. This type of adenocarcinoma is more frequent in older females and in proximal colon, although being also found in the left-side colon and mostly in the rectum. The majority of serrated adenocarcinomas appear in association with traditional serrated adenomas that usually are MSS or MSI-L. The tumors arising from sessile serrated polyps (15-20%) are frequently MSI-H and usually are right-sided [55, 58]. Several studies have shown that MSS or MSI-L serrated adenocarcinomas have a worse prognosis than non-serrated adenocarcinomas in the same conditions [59].

The genetic alterations observed in serrated pathway differ from the ones found in the adenoma-carcinoma pathway. Indeed, in the serrated pathway *APC* and *TP53* mutations and loss of heterozygosity are rare, whereas hypermethylation of CpG islands and alteration of microsatellite sequences are quite common. Right-side tumors are predominantly MSI-H, CIMP and *BRAF* mutant, while the left-side tumors are mainly MSS or MSI-L and are associated with *KRAS* mutations (Figure 8) [55]. In conclusion, there are two very important steps in serrated pathway: MAPK pathway activation and CIMP. The first occurs by either *BRAF* or *KRAS* mutation and the second can be either CIMP-low or CIMP-high. As mentioned before, MSI is quite frequent although is not a requirement of this pathway [58].

In CRC, *KRAS* and *BRAF* activating mutations are mutually exclusive, but. *PIK3CA* mutations can coexist with mutations in any of the former genes. The *KRAS* oncogenic activation mutations at codons 12, 13, or 61 have been reported to be followed by *APC* inactivation during tumor progression [9].

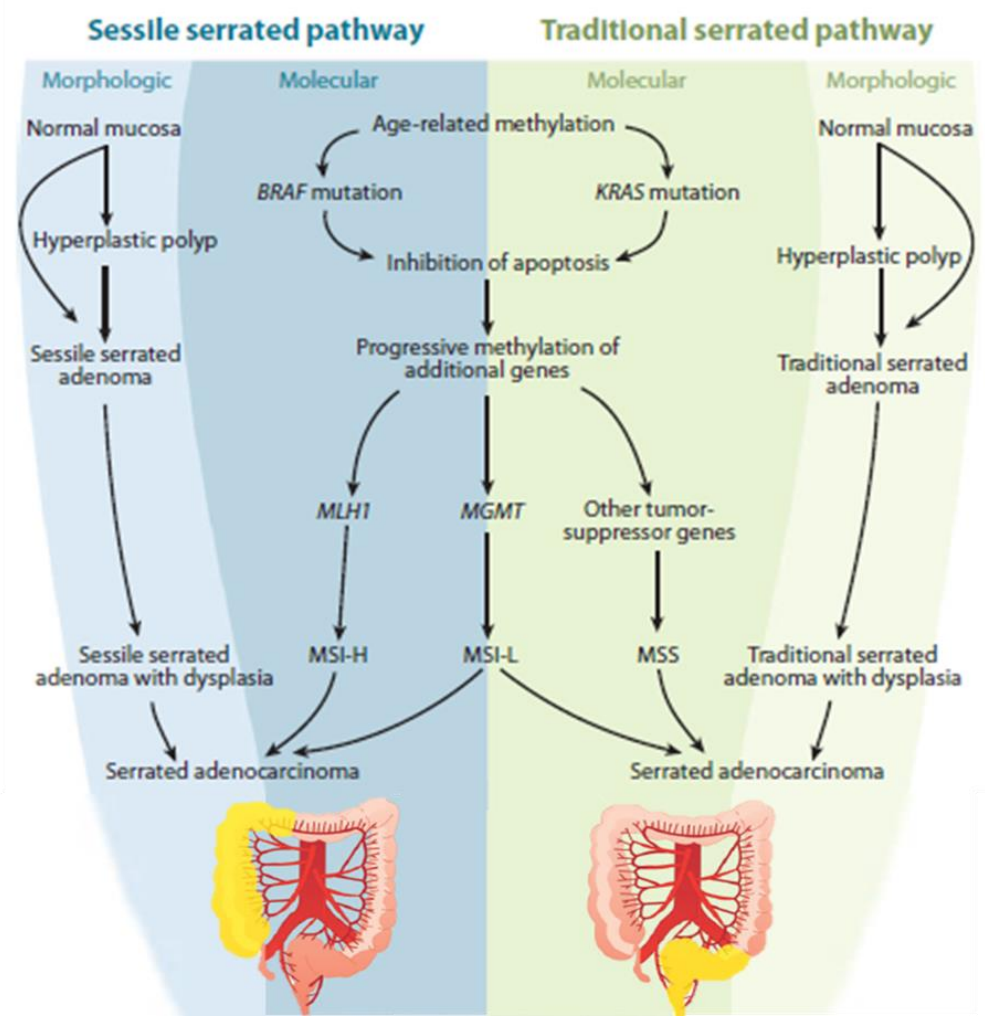

Figure 8 – Diagram of serrated neoplastic pathway with both morphologic and molecular steps involved, according to the adenoma type associated [55].

Biomarkers in CRC diagnosis and prognosis

A biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” by Biomarkers Definitions Working [60]. In addition, a biomarker should have the greatest value in early efficacy and safety evaluations [60].

DNA-based biomarkers have several advantages, being more stable than RNA and proteins, more practical, reliable and viable [34]. Specifically, for CRC, the search for biomarkers mostly based on genes promoter methylation seem to be a rather promising. A number of methods might be used to detect promoter methylation in tumors, including methylation specific PCR (MSP), real-time PCR (such as MethyLight) and

bisulfite pyrosequencing [61]. Remarkably, methylation might be assessed in cell free DNA extracted from body fluids like plasma, serum or stool [39].

Several studies have been made to discover a methylation biomarker or a panel of biomarkers with high sensitivity and specificity to be used in diagnosis and prognosis of CRC [46-48, 62]. However, only a few biomarkers have been tested the clinical trials and are commercially available. A recently licensed DNA methylation assay is *ColoVantage*®, which is blood-based and evaluates *septin 9* (*SEPT9*) methylation [62].The overall sensitivity of the *SEPT9* methylated DNA assay was found to be 90 %, with 87% sensitivity for early stage CRC (stages 1 and 2) and 100% at late stage diagnosis (stages 3 and 4) [62]. *Epi proColon*® 2.0 and Abbott RealTime mS9 are other assays based on *SEPT9* methylation that are currently used for CRC detection [63]. Nonetheless, the value of the *SEPT9* promoter methylation has been questioned. Indeed, a recent study found a standardized sensitivity of only 48.2 % for CRC detection, although high specificity (91.5%) was confirmed [64].

Another assay clinically available is *ColoSure*™ test, which is a fecal-based methylation assay that assess *vimentin* methylation [65]. This test is recommended for use along with colonoscopy since its sensitivity ranged from 38-88% [65, 66].

In addition to these DNA methylation biomarkers already being commercialized, there are others currently under investigation and/or clinical trials and the research for more effective potential biomarkers is ongoing [66]. For this reason, further studies are needed to find a more accurate methylation biomarker panel to be used in diagnosis and prognosis of CRC.

In order to find and select genes to be analyzed, a literature review was made. In **Appendix 1**, it is explained all the steps of this search in more detail. *APC* and *Ras association domain family 1 - isoform A* (*RASSF1A*) was reported to distinguish normal tissues from neoplastic whereas *MLH1*, *MGMT*, *SEPT9* and *insulin-like growth factor 2* (*IGF2*) were reported to specifically detect CRC.

APC is a tumor suppressor gene which encodes a protein involved in Wnt signaling pathway inhibition, therefore having an important role in cell-cycle regulation and apoptosis [32, 67]. *APC* promoter hypermethylation has been reported in several types of cancer, including colorectal [68, 69]; breast [70]; lung [71]; and prostate [72]; among others.

RASSF1A is also a tumor suppressor gene. The encoding protein of this gene acts as negative RAS effector, pro-apoptotic factor and it is involved in microtubule stabilization [32, 73]. Similarly to *APC*, this gene is reported to be hypermethylated in a number of cancers, including the four types mentioned for *APC* [73, 74].

MLH1 belongs to the mismatch repair system, consequently when it losses function or expression is lost, the cell is unable to repair DNA replication errors resulting in tumor initiation and progression [75, 76]. *MLH1* promoter hypermethylation has been described in sporadic colorectal cancer [69, 76] and in endometrial cancer [77].

IGF2 gene encodes a fetal growth factor which is normally imprinted. The *IGF2* loss of imprinting (LOI) has been associated with increased adenoma and colon cancer risk indicating its role on carcinogenesis [78]. Furthermore, *IGF2* LOI is correlated with hypomethylation at a differentially methylated region (DMR)-0 [79]. *IGF2* DMR-0 hypomethylation has been reported in colorectal cancer [78-80] and in breast cancer [81].

SEPT9 belongs to a highly conserved family of *septin* genes coding for GTP-Binding proteins that assemble into complexes forming filamentous structures of the cytoskeleton [82]. The septin proteins are important in many cellular processes, thus, having a major role in multiple cancers [82]. In addition of being methylated in colorectal cancer [62, 63, 82], it was also reported to be overexpressed in several types of tumors: breast, central nervous system, endometrium, kidney, liver, lung, lymphoid, esophagus, ovary, pancreas, skin, soft tissue and thyroid [83].

MGMT is a DNA repair gene which encodes a protein involved in defending cells against alkylating agents [84]. Promoter methylation associated with *MGMT* silencing have been reported mainly in colorectal cancer [85], but also in gliomas [86], gastric cancer [87], esophageal adenocarcinoma [88] and finally in oral squamous cell carcinoma [89].

Thus, the aberrant methylation of these genes may result in an accurate biomarker panel for diagnosis and prognosis of CRC.

Aims

In the present dissertation, the main goal was to develop a non-invasive method to detect primary colorectal tumors based on the evaluation of promoter methylation of a selected panel of genes.

Specifically, the aims of this study were:

- Assess the promoter methylation status of six selected genes (*APC*, *RASSF1A*, *MLH1*, *MGMT*, *IGF2* and *SEPT9*) in a clinically and pathologically well characterized cohort of CRC and normal tissue samples.
- Correlate the molecular results with standard clinicopathological parameters;
- Establish a panel of DNA-based methylation markers for the diagnosis of CRC and prognosis assessment.

Material and Methods

Study Group

A cohort of 214 samples of CRC and 50 samples of healthy controls (no evidence of CRC or other gastrointestinal cancer) were included. The CRC's tissue samples were all from a primary tumor and represent the population-based CRC data in order to be a relatively unbiased cohort. This project was approved by the institutional ethics committee (CES 120/015).

All the samples were obtained from formalin-fixed paraffin-embedded (FFPE) tissue blocks from the IPO-Porto hospital where the cohort of patients had undergone resections of primary tumors. For healthy controls, blocks from surgical margins of non-gastrointestinal tumors with total absence of tumor tissue of any kind were used. Information from relevant clinicopathological parameters from all patients was collected from the patients' clinical charts and put together in a data base to further analyses, like explained in **Appendix 2**.

Histopathologic evaluations

Haematoxylin and eosin (H&E) stained tissue sections of all the samples were examined under a light microscope by a pathologist to ensure that it was tumor or normal tissue, depending on the case. Also during the examination, the pathologist did a delimitation of the tumor area or normal mucosa depending on the sample. Tumors were classified into well, moderately or poorly differentiated.

Immunohistochemistry for *MLH1*, *MSH2*, *MSH6* and *PMS2* was performed by a technician for all of the CRC cases in order to evaluate any alteration in its expression and so understand which cases followed the MSI pathway, cases lacking expression of any of these proteins were classified as MSI-H, all the others were considered MSI-L or MSS. This evaluation was performed by a pathologist.

Genomic DNA extraction

DNA was extracted from FFPE tissue samples, using phenol-chloroform conventional method as described by Ramalho-Carvalho, Pires [86].

For each patient, one paraffin-embedded tissue block was considered and 12 serials 8 µm thick unstained slides and two H&E stained slides were made by a technician. As described before, the H&E stained slides were examined and the area of interest was delimitated. The two stained slides were made with the first and last section from the paraffin block, all the other slides were within these two, therefore it was possible to assess if the tumor's shape has change between the first and last section.

Before the DNA extraction, the unstained slides were incubated at 55 °C for about 30 minutes allowing the melting of some paraffin and the decrease of its amount around the tissue and, this way, facilitating the next step. Using a succession of reagents, the slides were first immerse in two xylene containers for 5 minutes in each, allowing the tissue deparaffination, and next, they were immerse in containers with 100%, 90%, 70% and 50% ethanol for 5 minutes in each by this order, allowing the tissue's rehydration.

Subsequently the tumor areas were delimited, by comparison with correspondent H&E stained slides, and macro-dissected from the tissue unstained slides. The macro-dissected tissue was resuspended in 1000 µL of digestion buffer (Tris-HCl 0,05M, EDTA 0,05M and Tween 20) and 25 µL of proteinase K (20 mg/mL) (NZYTech, Portugal) was added. Samples were incubated overnight at 55°C and after this period, if the sample was not totally digested, this incubation would be prolonged and more proteinase K would be added until complete digestion was achieved.

Extraction was performed with phenol-chloroform (Sigma-Aldrich®, Germany; Merck, Germany) using Phase Lock Gel™ Light 2 mL tubes (5PRIME, Germany). Briefly, the samples were transferred to the phase lock tubes, previously centrifuged at 12,000 rpm for 5 minutes, and mixed with 500 µL of phenol-chloroform. After centrifugation at 13,000 rpm for 15 minutes, the aqueous phase containing the DNA (upper phase) was transferred to 2 mL tubes.

DNA precipitation was accomplished by adding to the transferred aqueous phase, chilled absolute ethanol (2 volumes of the original amount of this phase) (Merck Millipore, Germany), 7.5M ammonium acetate (1/3 volume) (Sigma-Aldrich®, Germany) and 2µL of glycogen (Thermo Fisher Scientific, USA). After mixing, samples were left at -20°C overnight.

In the next step, DNA was washed twice with 70% ethanol using centrifugations at 13,000 rpm for 20 minutes. The pellets were air dried and eluted in 20µL of sterile distilled water (B.Braun, Melsungen, Germany). DNA was quantified and its purity assessed using NanoDrop ND-1000 ® (NanoDrop Technologies, DE, USA) spectrophotometer. All the samples were stored at -20°C until further use.

Sodium Bisulfite Treatment

Sodium bisulfite treatment was performed in order to be able to study the methylation status. This method is based on a differential sensitivity to chemical conversion, where unmethylated cytosine residues are converted into uracil, while 5-methylcytosines remain unchanged (Figure 9). Thus, an epigenetic event is converted into a genetic change allowing its analysis by PCR-based methods [90].

This method includes several steps: starting with the DNA denaturation to improve the action of bisulfite; a sulphonation resulting in a cytosine sulphonate derivative; a hydrolytic deamination of the previous compound leading to the formation of a sulphonated uracil derivative; and at last, the removal of the sulphonate group by a subsequent alkali treatment originating the uracil residue [90].

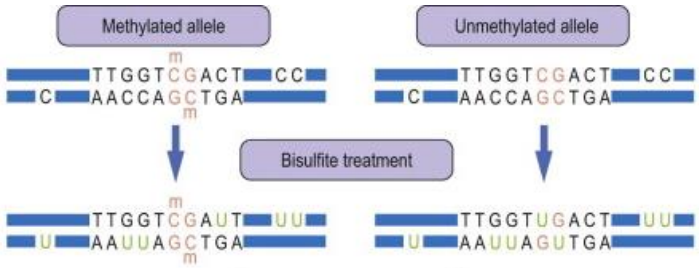

Figure 9 – Changes in DNA caused by bisulfite treatment and the differences between methylated CpG sites and unmethylated ones (Adapted from [91]).

For each sample, 1000 ng of DNA was sodium bisulphite modified using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's instructions.

Briefly, the sample volume containing 1000 ng of DNA was diluted in sterile water up to 20 µL of total volume in a PCR tube. To each tube was added 130 µL of CT Conversion Reagent and then incubated in a GeneAmp PCR System 2700 thermal cycler (Applied Biosystems, USA) at 98°C for 10 minutes and then at 64°C for 3 cycles of 60 minutes each. After, it can be at 4°C up to eighteen hours.

Samples were transferred to a Zymo-Spin IC column already containing 600 µL of M-binding buffer and centrifuged at 10,000 rpm for 30 seconds. Next, the column was washed with 100 µL of M-Wash buffer and once again centrifuged at 10,000 rpm for 30 seconds. The column was incubated at room temperature with 200 µL of M-Desulphonation buffer for 20 minutes in order to desulphonate the sample. After the incubation, the columns were centrifuged at 10,000 rpm for 30 seconds followed by two washing steps with 200 µL of M-Wash buffer and centrifugations at 10,000 rpm for 30

seconds. One last centrifugation at 10,000 rpm for 30 seconds was performed to completely dry the column.

The column was removed from the collection tube and placed in a 1.5 mL tube. The modified DNA was eluted by incubating the column with 30 µL of sterile distilled water for 5 minutes at room temperature followed by a centrifugation at 12,000 rpm for 30 seconds. This last step was repeated allowing a final volume of 60 µL of modified DNA for each sample. Bisulfite modified DNA was stored at -80°C until further use.

CpGenome™ Universal Methylated DNA (Merck Millipore, Germany) was also modified, using the guidelines described above and eluted in a total of 20 µL of sterile distilled water.

Quantitative DNA Methylation analysis using qMSP

The methylation levels of the selected genes were assessed using quantitative methylation specific PCR (qMSP). The modified DNA was used as template and a set of primers designed specifically for bisulfite-converted DNA were used for each one of the target genes: *APC*, *IGF2*, *MGMT*, *MLH1*, *RASSF1A* and *SEPT9* (Table 7).

Quantitative methylation specific PCR is a highly sensitive and specific method that when compared to conventional PCR assays, allows an accurate and high sensitive DNA concentration determination [92]. Moreover, the obtain results can be either qualitative or quantitative, showing the present or absence of the DNA sequence (Figure 10) of interest and when it is present, gives a relative quantification [92].

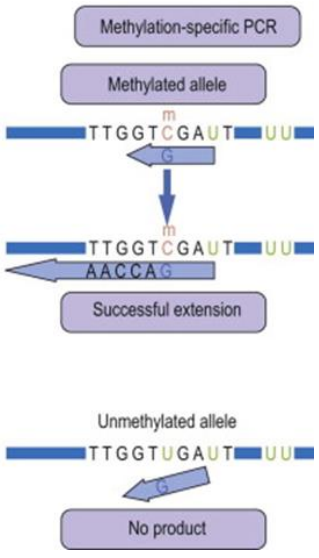

Figure 10 – Scheme explaining methylation-specific PCR in the presence or absence of methylation using the appropriate primers (Adapted from [91]).

For the amplicon detection was used a double-stranded (ds) DNA-intercalating agent (SYBR Green I). By this method, the intensity of the fluorescence signal is therefore dependent on the quantity of dsDNA present in the reaction. The main advantage of this detection method is being relatively cheap and that can be used with any pair of primers for any target, and the disadvantage is not being specific, since the dye binds to all dsDNAs formed during the PCR reaction [92, 93]. To ensure that the fluorescence observed was specific and not due to cross dimer of primers, the melting curve was also evaluated.

The quantification depends on the DNA input, and for this reason, to estimate the input DNA quantity, the control gene *ACTβ* was used for normalization [94].

Modified CpGenome™ Universal Methylated DNA was used to create five serial dilutions by a 5x dilution factor. These serial dilutions were used to generate a standard curve allowing absolute quantification as well as ascertaining PCR efficiency.

The reactions were carried out in 384-well plates (4titude, UK) using LightCycler 480 II (Roche, Germany). Two µL of modified DNA, 5 µL of KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems, USA), 0.4 µL (*ACTβ*) or 0.3 µL (target gene) 10 µM primers (F+R) and 2.6 µL or 2,7 µL of sterile distilled water, were added per well. All the samples were run in triplicates while the standards and negative control were run in duplicates. The primer sequences used for each gene are displayed on Table 7.

Table 7 – Primers sequence used and qMSP condition for each one of the target genes.

Gene	Forward (F) 5'-3'	Reverse (R) 5'-3'	Annealing Temperature (°C)	Concentration (µM)
ACTβ	TGGTGATGGAGGAGGTTTAGTAAGT	AACCAATAAAACCTACTCCTCCCTTAA	60	0.4
APC	TGTGTTTTATTGCGGAGTGC	CACATATCGATCACGTACGC	62	0.3
IGF2	CGTCGTTTTTTATTGGTTTC	CACACGAATAACCCGCCT	60	0.3
MGMT	TTTCGACGTTCGTAGGTTTTCGC	GCACTCTTCCGAAAACGAAACG	60	0.3
MLH1	CGTTATATATCGTTCGTAGTATTCGTGTTT	CTATCGCCGCCTCATCGT	62	0.3
RASSF1A	AGCGAAGTACGGGTTTAATC	ACACGCTCCAACCGAATA	60	0.3
SEPT9	TTAGTTAGCGCGTAGGGTTC	ACCTTCGAAATCCGAAATAA	60	0.3

The PCR program consisted in 1 cycle at 95°C for 3 minutes (for enzyme activation), followed by 45 cycles at 95°C with 3 seconds each (for DNA denaturation) and 1 last cycle at 60°C or 62°C (see Table 7) for 30 seconds (for annealing, extension

and data acquisition). In the end, the program assesses the melting curve of all the products formed during the PCR reaction.

All the plates showed 95-100% efficiency and all the standard curves had a slope within -3.33 and -3.92. The plates from the same gene had equal efficiency or the difference between them was less than 1%. For the samples, the standard deviation was always less than 0.38, considering at least two of the three replicates.

Considering the results obtain from qMSP, the Relative Methylation Levels, for each sample, were calculated as the ratio between the target gene mean quantity and the reference gene mean quantity, multiplied by one thousand.

Statistical analysis

Statistical analysis was performed using SPSS Statistics 22 (IBM, USA).

The gene methylation levels and clinical parameters were compared within groups using non-parametric tests. It was used Mann-Whitney U test for comparisons between two groups and Kruskall-Wallis test for comparisons between three or more groups. A *P*-value <0.05 was considered statistically significant.

For diagnosis analysis were only considered the 145 CRC cases that didn't had any treatment prior to surgery. For most of the analysis the cases were divided by tumor location, colon (proximal and distal) and rectum.

To evaluate the biomarkers performance in terms of diagnosis, ROC curves were built for each one of them. Next, parameters as specificity and sensitivity were determined for the individual biomarkers. Forward and backward stepwise elimination using a binary logistic regression was performed to select biomarkers in the final model. A ROC curve for the final model was built and the correspondent specificity and sensitivity was calculated. For this last task, the panel was considered as positive for a specific sample when at least one of the genes were positive in the individual model.

Still for diagnosis evaluation, the positive predictive value (PPV) and negative predictive value (NPV) were computed according to the following formulation: PPV = true positives/all positives and NPV = true negatives/all negatives.

The disease-free survival (DFS) and disease-specific survival (DSS) analysis was performed in order to assess the biomarkers' prognosis value. At first, it was used the Kaplan-Meier method, using the two-sided log-rank test to compare the survival curves. Then, the Cox model was fitted to assess the prognostic value of the clinical parameters and methylation status of the selected genes. Univariable test was used to assess the contribution of each variable by determining their Hazard ratios (HR), followed

by a multivariable test considering a group of clinical parameters. For the multivariable test, all the variables with a significant association, with DFS or DSS, in univariable analysis were included.

For the prognosis analysis, the methylation levels were divided by 25th percentile for *IGF2* and 75th percentile for all the other genes. For *IGF2*, methylation levels under the percentile were called hypomethylated and above it, by non-hypomethylated. For all the remaining genes (*MGMT, MLH1, RASSF1A, SEPT9*), methylation levels above the percentile were considered hypermethylated and non-hypermethylated, when inferior. The non-hypermethylated was considered the reference group.

Results

Clinical dataset

A total of 214 patients, diagnosed with CRC between 2000 and 2012, with the exception of two cases, one from 1994 and another from 1997, were included in this study, except for one patient that was lost to follow-up. The cohort comprised: 110 colon cancer patients, and 104 rectum cancer patients (Table 8 and Table 9). The relevant clinical information for the 50 normal samples are displayed in Table 10.

Table 8 – Clinicopathologic features of CRC patients by tumor location.

Characteristic	Total (n = 214)	Colon (n = 110)	Rectum (n = 104)
Age (years)			
Mean	60.35 ± 0.66	60.82 ± 0.973	59.80 ± 0.898
Range	25-80	25-80	31-80
Gender			
Female	74 (34.6%)	42 (38.2%)	32 (30.8%)
Male	140 (65.4%)	68 (61.8%)	72 (69.2%)
Stage			
I and II	52 (24.3%)	24 (21.8%)	28 (26.9%)
III	52 (24.3%)	20 (18.2%)	32 (30.8%)
IV	108 (50.5%)	64 (58.2%)	44 (42.3%)
Unknown	2 (0.9%)	2 (1.8%)	-
Tumor differentiation			
Well	4 (1.9%)	1 (0.9%)	3 (2.9%)
Moderate	123 (57.5%)	79 (71.8%)	44 (42.3%)
Poor	5 (2.3%)	4 (3.6%)	1 (1.0%)
Not Assessable	82 (38.3%)	26 (23.6%)	56 (53.8%)
Histology			
Mucinous	16 (7.5%)	12 (10.9%)	4 (3.8%)
Other	198 (92.5%)	98 (89.1%)	100 (96.2%)
KRAS mutation			
Wildtype	116 (54.2%)	60 (54.5%)	56 (53.85%)
Mutated	84 (39.3%)	46 (41.8%)	38 (36.54%)
Was not assessed	14 (6.5%)	4 (3.6%)	10 (9.61%)
MSI			
MSI-H	8 (3.7%)	8 (7.3%)	0 (0%)
MSI-L or MSS	206 (96.3%)	102 (92.7%)	104 (100%)
Neoadjuvant treatment			
Yes	69 (32.2%)	16 (14.5%)	53 (51.0%)
No	145 (67.8%)	94 (85.5%)	51 (49.0%)
Adjuvant treatment			
Yes	171 (80.0%)	89 (80.9%)	82 (78.8%)
No	45 (20.0%)	21 (19.1%)	22 (21.2%)

Table 9 – Number of cases by tumor location: rectum, distal or proximal colon.

Tumor Location	Total (n = 214)
Proximal colon	37 (17.3%)
Distal colon	73 (34.1%)
Rectum	104 (48.6%)

Table 10 – Basic features of Normal colon/rectum samples used.

Characteristic	CRN (n = 50)
Age (years)	
Mean	57.50 ± 2.15
Range	18-85
Gender	
Female	32 (64.0%)
Male	18 (36.0%)
Tissue Location	
Proximal colon	19 (38.0%)
Distal colon	16 (32.0%)
Rectum	6 (12.0%)
Not Assessable	9 (18.0%)

Assessment in clinical samples

The selected genes' methylation levels were assessed for all the CRC and CRN samples using qMSP. For *SEPT9*, *RASSF1A*, *MGMT* and *APC*, relative methylation levels were significantly higher in CRC compared to normal tissues (Figure 11), whereas *IGF2* relative methylation levels were significantly lower in CRC. Specifically, by location, *SEPT9*, *RASSF1A* and *MGMT* methylation levels were significantly higher in colon cancer patients, whereas *SEPT9*, *MGMT* and *APC* were significantly hypermethylated in rectum cancer, compared to CRN (Figure 12). Similar results to colon were observed when a separated analysis of proximal and distal colon was performed (Mann-Whitney U Test, $P<0.05$).

Moreover, slightly higher *SEPT9* methylation levels were found in colon cancer patients than in rectum cancer patients (Mann-Whitney U Test, $P=0.021$). No significant association was found between genes' methylation levels and patients' age.

MGMT and *MLH1 m*ethylation levels were marginally, but significantly higher in female patients compared with males both in colon and rectum samples ($P=0.048$ and $P=0.049$, $P=0.007$ and $P=0.010$, respectively).

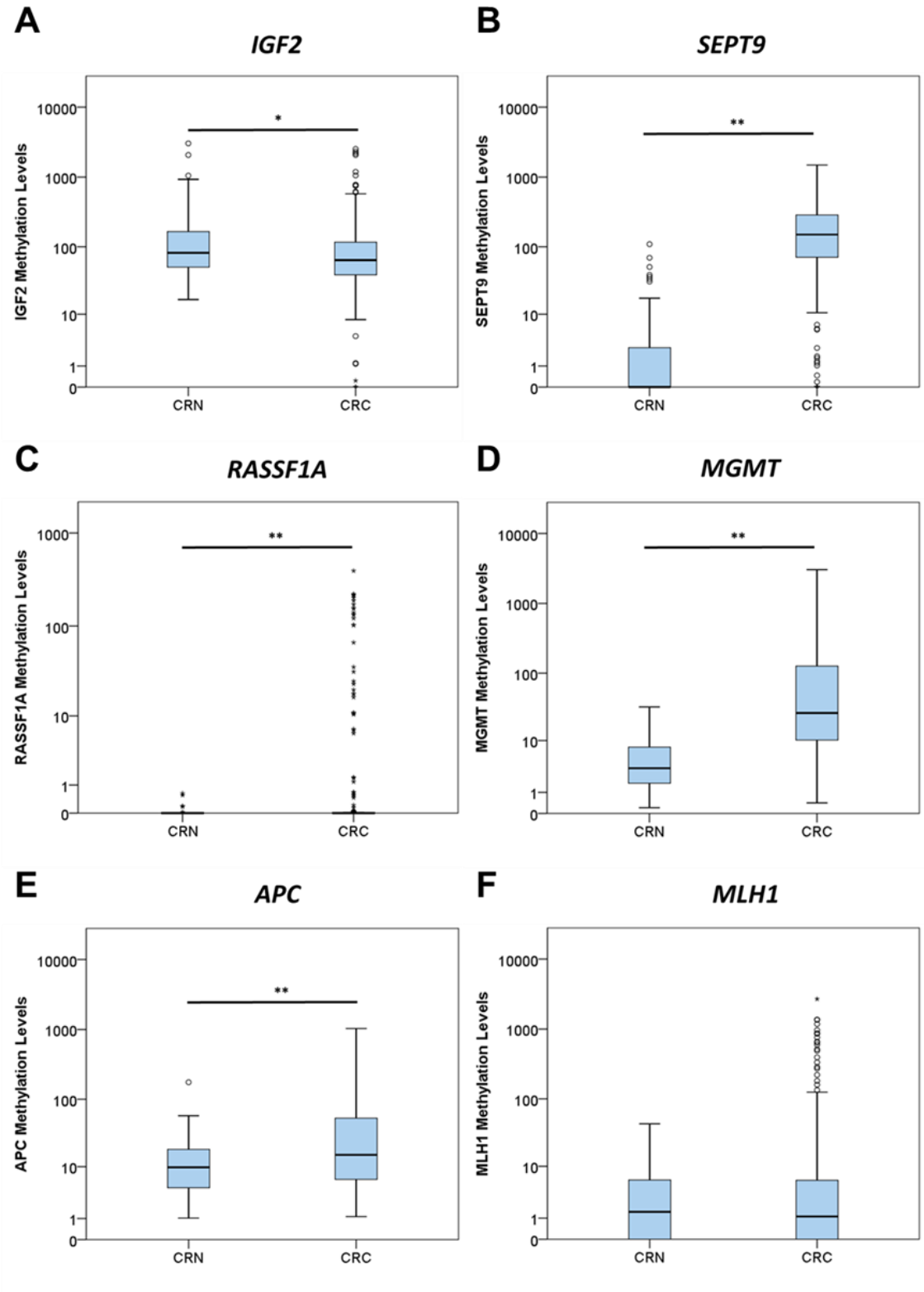

Figure 11 – Box-plots of (A) *IGF2*, (B) *SEPT9*, (C) *RASSF1A*, (D) *MGMT*, (E) *APC*, (F) *MLH1* promoter methylation levels in normal (CRN) and neoplastic tissue (CRC). (Mann-Whitney U Test, * *P*<0.05; ** *P*<0.01).

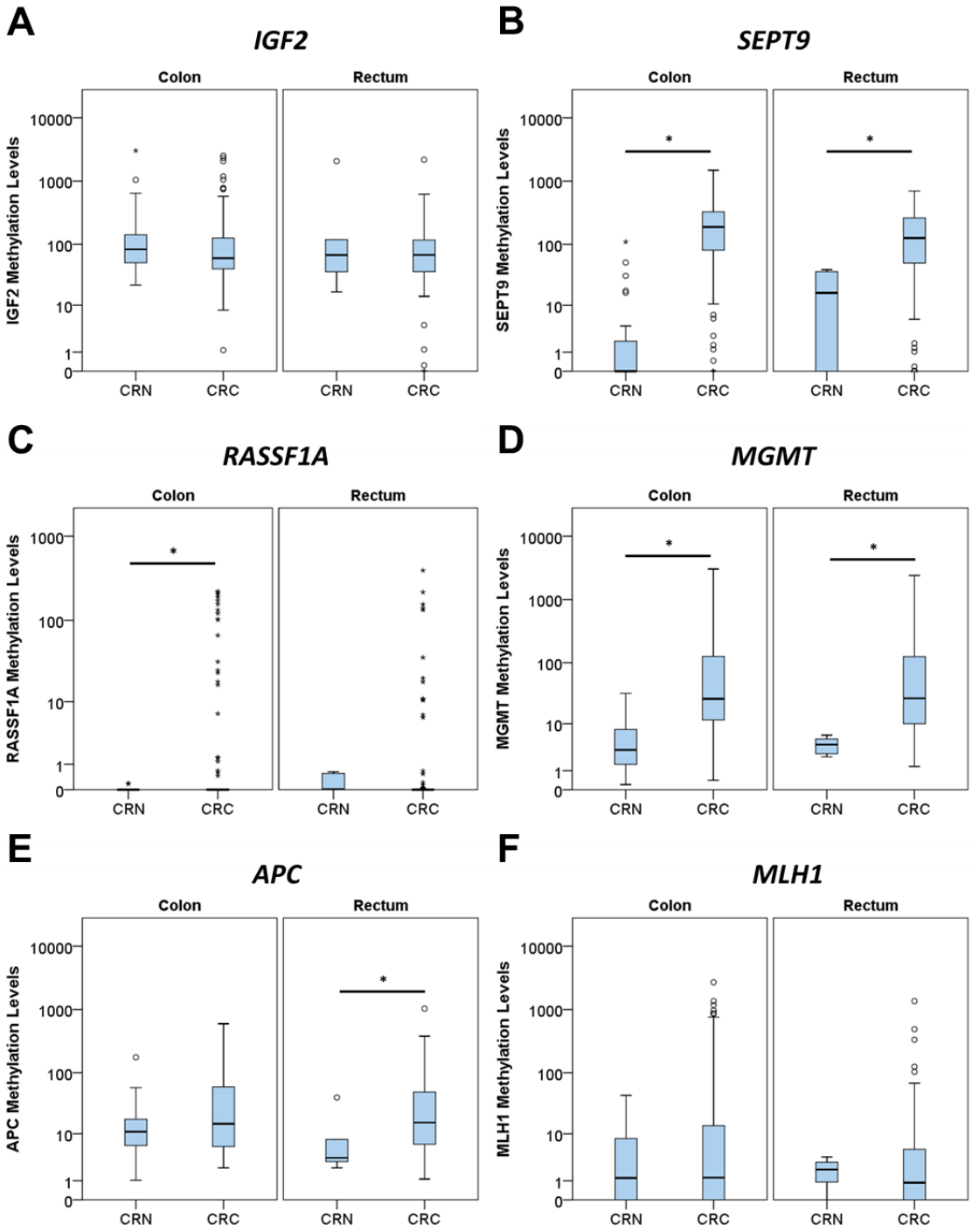

Figure 12 – Box-plots for (A) *IGF2*, (B) *SEPT9*, (C) *RASSF1A*, (D) *MGMT*, (E) *APC*, (F) *MLH1* promoter methylation levels in normal (CRN) and neoplastic tissue (CRC), by location (Mann-Whitney U Test, * *P*<0.01).

The methylation levels for each gene were compared for the most relevant features considering colon and rectum samples in separated (**Appendix 3**: Table 22 and Table 23). Neoadjuvant treated rectum cancer patients showed significantly lower *SEPT9* and *MGMT* methylation levels (*P*=0.002 and *P*=0.012, respectively).

MSI *status* assessed by immunohistochemistry showed that 8 cases were MSI-H, whereas 206 were MSI-L or MSS. Except for *APC* promoter (P=0.008), no significant differences were found for methylation levels between MSI-H and MSI-L/MSS (Table 11, Figure 13).

Table 11 – Comparative analysis of genes methylation levels according to the MSI status (MSI-H or MSI-L/MSS).

Gene	*IGF2*	*SEPT9*	*RASSF1A*	*MGMT*	*APC*	*MLH1*
P	0.141	0.816	0.080	0.701	**0.008**	0.368
Test	Mann-Whitney U Test					

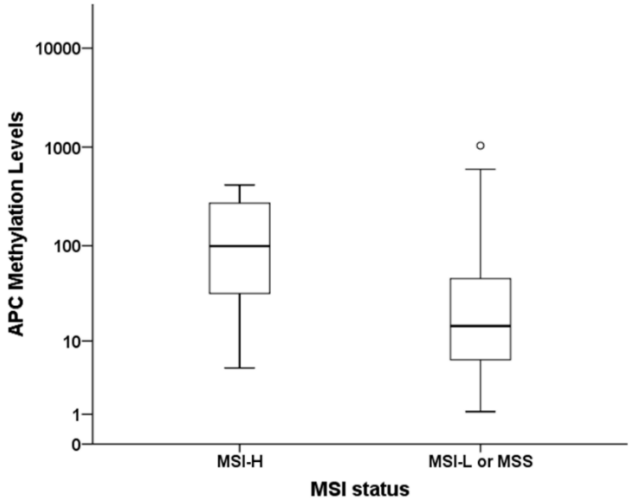

Figure 13 – Box-plot of *APC* Methylation levels across MSI-H and MSI-L or MSS samples. Mann-Whitney U Test, *P*=0.008.

Evaluation of the diagnostic value

The best performance for CRC detection was achieved by *SEPT9*, with 85.5% sensitivity, 94.0% specificity, 97.6% PPV and 69.1% NPV.

Table 12 – Performance of epigenetic biomarkers for detection of CRC in tissue. Sensitivity and specificity values (%) for the individual genes and panels. n stand for number of positive cases and N for the total of cases tested.

	Sensitivity (n/N)	Specificity (n/N)	AUC
SEPT9	85.5 (124/145)	94.0 (3/50)	0.950
MGMT	77.2 (112/145)	84.0 (8/50)	0.894
RASSF1A	33.1 (48/145)	90.0 (5/50)	0.621
IGF2 - hypomethylation	58.6 (85/145)	64.0 (18/50)	0.599
APC	35.2 (51/145)	90.0 (5/50)	0.651
MLH1	13.8 (20/145)	97.9 (1/48)	0.496
Panel			
SEPT9, MGMT	93.8 (136/145)	82.0 (9/50)	0.964
SEPT9, MGMT, RASSF1A	96.6 (140/145)	74.0 (13/50)	0.970
SEPT9, MGMT, APC	95.2 (138/145)	76.0 (12/50)	0.965

Considering gene panels, the best performance was attained for *SEPT9/MGMT/RASSF1A*, with 96.6% sensitivity and 74.0% specificity (Figure 14). The highest values of PPV and NPV were obtained with *SEPT9/MGMT/APC* panel (92.0% and 87.3%, respectively) (Table 12, Table 13).

Table 13 – Positive and negative predictive values for *SEPT9* alone and for the three more relevant panels.

Panel	PPV (%)	NPV (%)
SEPT9	97.6	69.1
SEPT9, MGMT	93.8	82.0
SEPT9, MGMT, RASSF1A	91.5	72.5
SEPT9, MGMT, APC	92.0	87.3

ROC curve analysis was performed for all the selected genes, using only the cases without any treatment prior to surgery (n=145). The best performance was found for *SEPT9* followed by *MGMT* (AUCs of 0.950 and 0.894, respectively). Considering the gene panels, *SEPT9/MGMT/RASSF1A* methylation levels disclosed the best performance with 0.970 AUC.

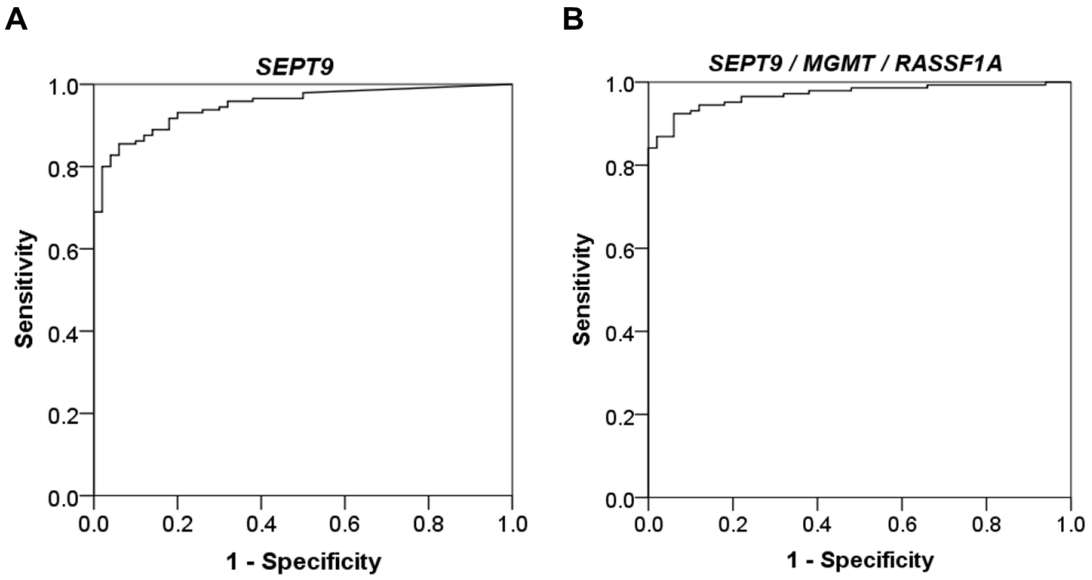

Figure 14 – ROC curves for promoter methylation of *SEPT9* alone and SEPT9/*MGMT*/*RASSF1A* panel considering the CRC cases without any treatment prior to surgery.

The positive detection rate was calculated according with lesion locations and CRC stages. Overall, the panel *SEPT9, MGMT, RASSF1A* performed better to detect cancer in all locations and stages of the disease (Table 14).

Table 14 – Positive detection rate for location groups and stage of disease. n stand for number of positive cases and N for the total of cases tested. Considering 145 cases in total.

Positive detection Rate [%(n/N)]			
Diagnosis Group	***SEPT9***	***SEPT9, MGMT, RASSF1A***	***SEPT9, MGMT, APC***
Location			
Colon:	84.0 (79/94)	95.7 (90/94)	93.6 (88/94)
Proximal	86.7 (26/30)	93.3 (28/30)	93.3 (28/30)
Distal	82.8 (53/64)	96.9 (62/64)	93.8 (60/64)
Rectum	88.2 (45/51)	98.0 (50/51)	98.0 (50/51)
Stage			
Stage I and II	91.4 (32/35)	100 (35/35)	97.1 (34/35)
Stage III	80.0 (28/35)	94.2 (33/35)	91.4 (32/35)
Stage IV	84.9 (62/73)	95.9 (70/73)	95.9 (70/73)

Evaluation of Prognostic Value

To define the prognostic value of selected genes' promoter methylation in CRC, disease-free survival (DFS) and disease-specific survival (DSS) were computed based on clinical and epigenetic variables (Appendix 4: Table 24 and Table 25). The median follow-up in this cohort was 4.32 years (range from 0.41 to 17.68 years). At the time of the last follow-up, 187 patients (87.4%) deceased from CRC, 12 (5.6%) are still being treated for CRC and 15 (7.0%) are considered disease-free.

Regarding DFS, none of the genes showed significant association with prognosis. As for the DSS, higher *SEPT9* and *MLH1* methylation levels were significantly associated with better prognosis, *P*=0.054 and *P*=0.017, respectively (Figure 15).

A multivariable analysis, using Cox regression, was also performed to assess the real influence of clinicopathological and epigenetic features in predicting DFS and DSS (Table 15 and Table 16). Hypermethylation of *SEPT9* and *MLH1* was associated with better prognosis, while larger primary tumor and age were shown to be independent prognostic factors for DFS.

Higher *SEPT9* and *MLH1* methylation levels independently predicted improved DFS (*P*=0.031 and *P*=0.027, respectively) and DSS (*P*=0.031 and *P*=0.033, respectively) (Figure 16 and Figure 17). Primary tumor (pT), regional lymph nodes (pN) and age were also shown to independently predict DSS.

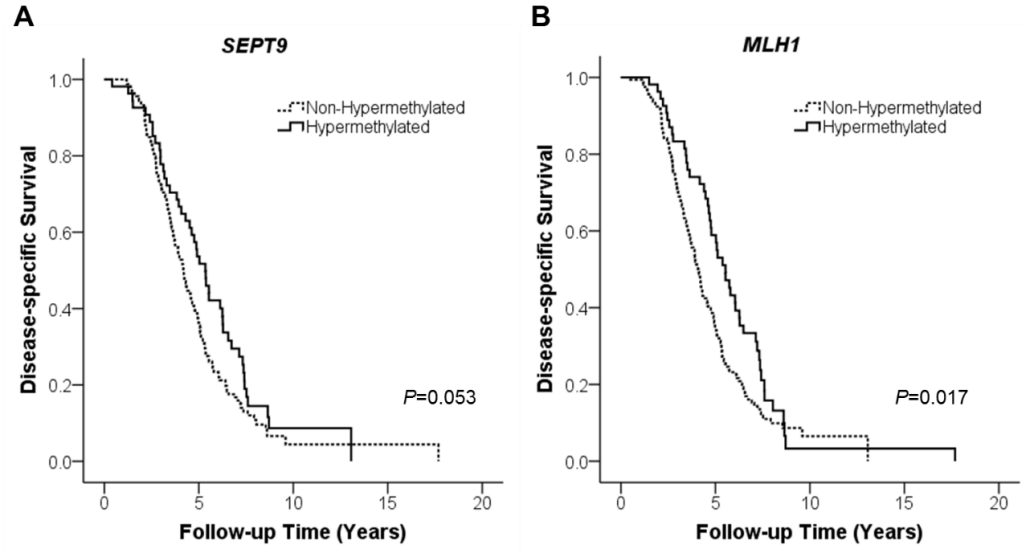

Figure 15 – Kaplan-Meier analysis with the association between DSS and methylation levels in *SEPT9* (A) and *MLH1* (B). *P* was assessed by Log-Rank Test. (n=214)

Table 15 – Multivariable analysis using Cox regression for Disease-free survival (DFS) (n=214).

Multivariable analysis		HR (95% CI)	*P*
SEPT9	Hyper	0.603 (0.381-0.954)	**0.031**
MLH1	Hyper	0.604 (0.386-0.945)	**0.027**
Primary Tumor (pT)	T1 and T2	1 (referent)	**<0.001**
	T3	1.664 (0.908-3.049)	0.099
	T4a and T4b	13.253 (3.708-47.364)	**<0.001**
Age group	18-57	1 (referent)	**0.018**
	58-65	1.993 (1.219-3.259)	**0.006**
	66-85	1.704 (0.997-2.912)	0.051

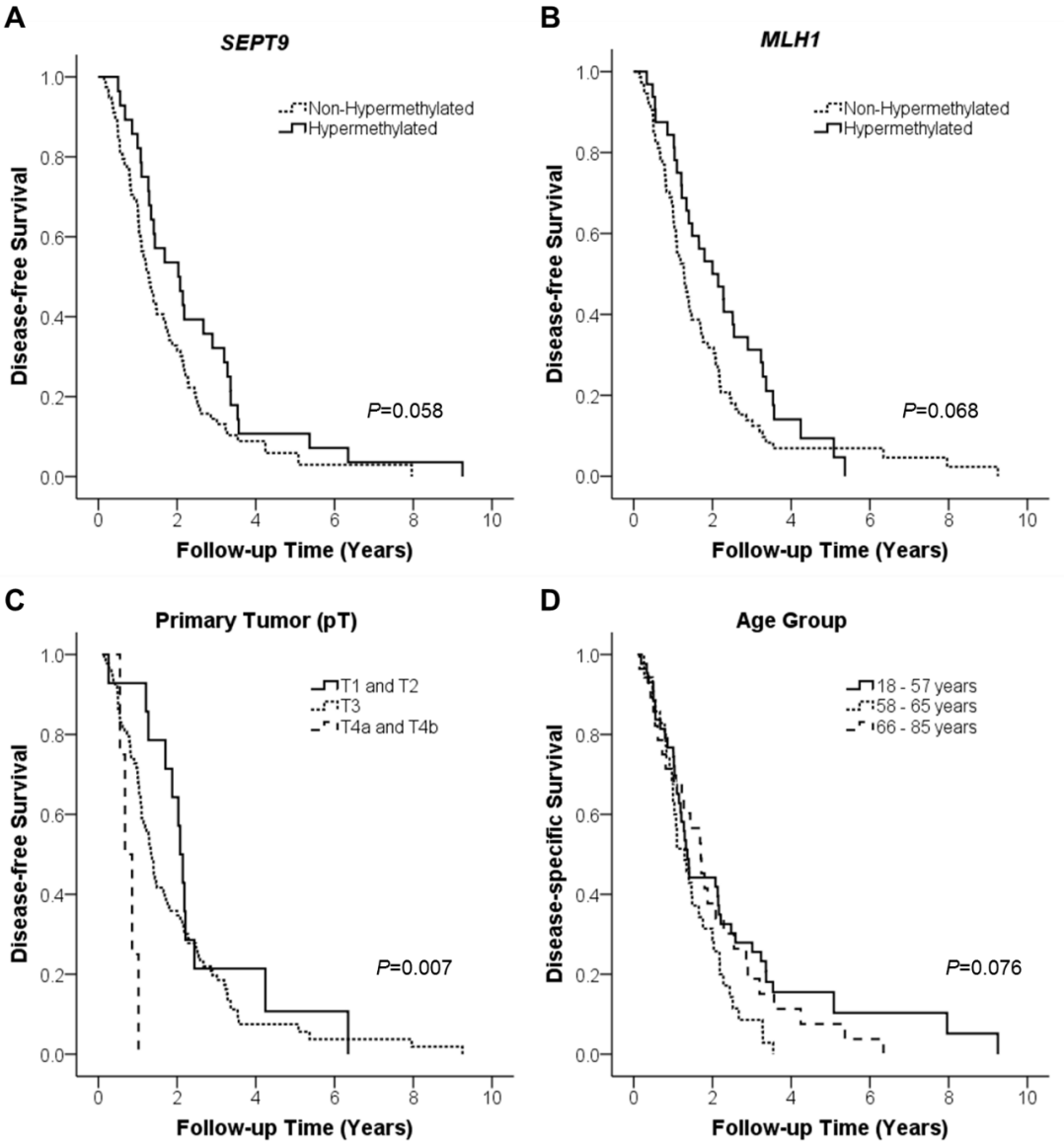

Figure 16 – Kaplan-Meier analysis with the association between DFS and SEPT9 (A) and MLH1 (B) methylation levels, pathological stage (C) and Age (D). *P* was assessed by Log-Rank Test. (n=214)

Table 16 – Multivariable analysis using Cox regression to assess the potential of *SEPT9* and *MLH1* methylation levels, in addiction to clinical features in prediction of Disease-specific survival (DSS). n=214.

Multivariable analysis		HR (95% CI)	*P*
SEPT9	Hyper	0.654 (0.445-0.962)	**0.031**
MLH1	Hyper	0.672 (0.467-0.968)	**0.033**
Stage	I and II	1 (referent)	0.106
	III	0.813 (0.423-1.563)	0.535
	IV	1.567 (0.795-3.087)	0.195
Primary Tumor (pT)	T1 and T2	1 (referent)	**0.005**
	T3	0.813 (0.496-1.334)	0.413
	T4a and T4b	2.189 (1.054-4.545)	**0.036**
Regional Lymph Nodes (pN)	N0	1 (referent)	**0.015**
	N1 a, b, c	0.907 (0.539-1.528)	0.714
	N2 a,b	1.798 (1.029-3.139)	**0.039**
Distant Metastasis (M)	M0	1 (referent)	0.680
	M1a	1.121 (0.652-1.926)	0.680
	M1b	-	-
Neoadjuvant Treatment	No	1 (referent)	-
	Yes	1.283 (0.904-1.821)	0.163
Age group	18-57	1 (referent)	**<0.001**
	58-65	2.035 (1.374-3.016)	**<0.001**
	66-85	2.297 (1.548-3.407)	**<0.001**
Colon	Rectum	1 (referent)	0.151
	Distal	0.969 (0.668-1.406)	0.868
	Proximal	1.504 (0.939-2.409)	0.090

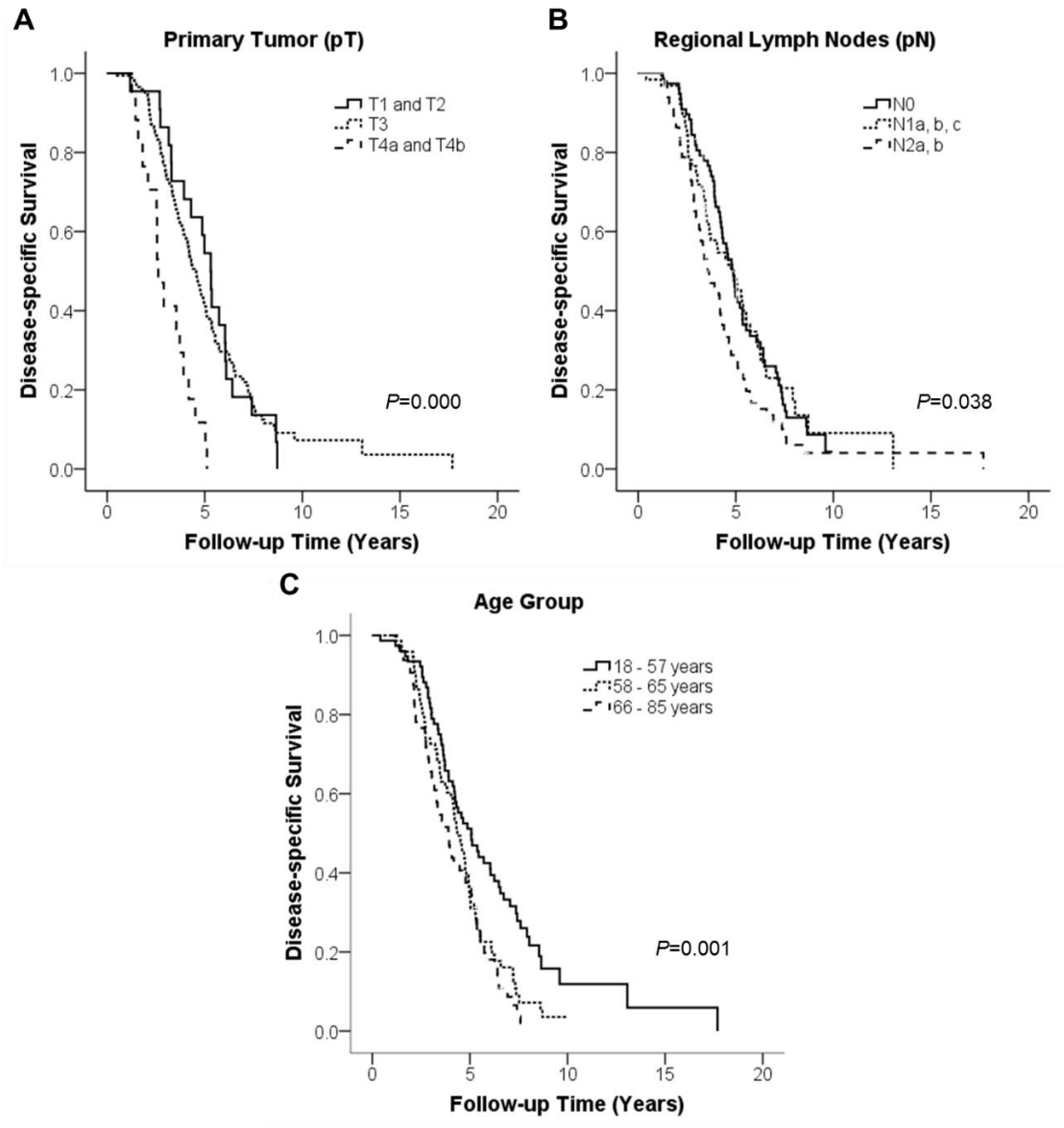

Figure 17 – Kaplan-Meier analysis with the association between DSS with primary tumor (A), regional lymph nodes (B) and age group (C). *P* was assessed by Log-Rank Test. (n=214)

To evaluate whether there was any specific association between selected genes methylation and tumor location (**Appendix 4**: Table 26, Table 27, Table 28, Table 29), univariable and multivariable analysis of DFS and DSS were also constructed for Colon and Rectum cancer independently (Table 17 and Table 18).

Aberrant methylation was not associated with DFS for any of the defined tumors' locations. Nonetheless, pT4 colon tumors associated with a shorter DFS (HR=11.44, 95% CI: 1.947- 67.229), whereas in rectum tumors, female gender associated with longer DFS (HR=0.296, 95% CI: 0.149 - 0.589).

Table 17 – Multivariable analysis using Cox regression to assess the potential of *SEPT9* and *MLH1* methylation levels and of clinical features in Disease-specific survival (DSS) in colon cancer patients (n=110).

Multivariable analysis		Colon	
		HR (95% CI)	***P***
SEPT9	Hyper	0.472 (0.275-0.809)	**0.006**
MLH1	Hyper	0.513 (0.302-0.873)	**0.014**
Stage	I and II	1 (referent)	0.720
	III	0.824 (0.396-1.716)	0.605
	IV	1.135 (0.517-2.490)	0.753
Primary Tumor (pT)	T1 and T2	1 (referent)	**0.001**
	T3	1.334 (0.495-3.594)	0.568
	T4a and T4b	4.643 (1.454-14.828)	**0.010**
Distant Metastasis (M)	M0	1 (referent)	-
	M1a	1.251 (0.633-2.470)	0.520
	M1b	-	-
Neoadjuvant Treatment	No	1 (referent)	-
	Yes	1.083 (0.593-1.976)	0.796
Age group	18-57	1 (referent)	**0.014**
	58-65	1.976 (1.071-3.647)	**0.029**
	66-85	2.258 (1.293-3.944)	**0.004**
Colon	Proximal	1 (referent)	**-**
	Distal	0.554 (0.345-0.891)	**0.015**

Regarding, DSS, in colon cancer, *SEPT9* and *MLH1* methylation levels significantly associated with better prognosis, whereas in rectum significance was only reached for *SEPT9* methylation.

Table 18 – Multivariable analysis using Cox regression to assess the potential of *SEPT9* and *MLH1* methylation levels, in addiction to clinical features in prediction of Disease-specific survival (DSS) on rectum. n=104.

Multivariable analysis (DSS)		Rectum	
		HR (95% CI)	***P***
SEPT9	Hyper	0.564 (0.319-0.996)	**0.049**
MLH1	Hyper	1.044 (0.615-1.773)	0.872
Stage	I and II	1 (referent)	0.124
	III	1.120 (0.632-1.984)	0.699
	IV	2.705 (1.032-7.091)	**0.043**
Distant Metastasis (M)	M0	1 (referent)	-
	M1a	0.813 (0.325-2.034)	0.658
	M1b	-	-
Age group	18-57	1 (referent)	**0.040**
	58-65	1.889 (1.100-3.245)	**0.021**
	66-85	1.848 (1.045-3.268)	**0.035**

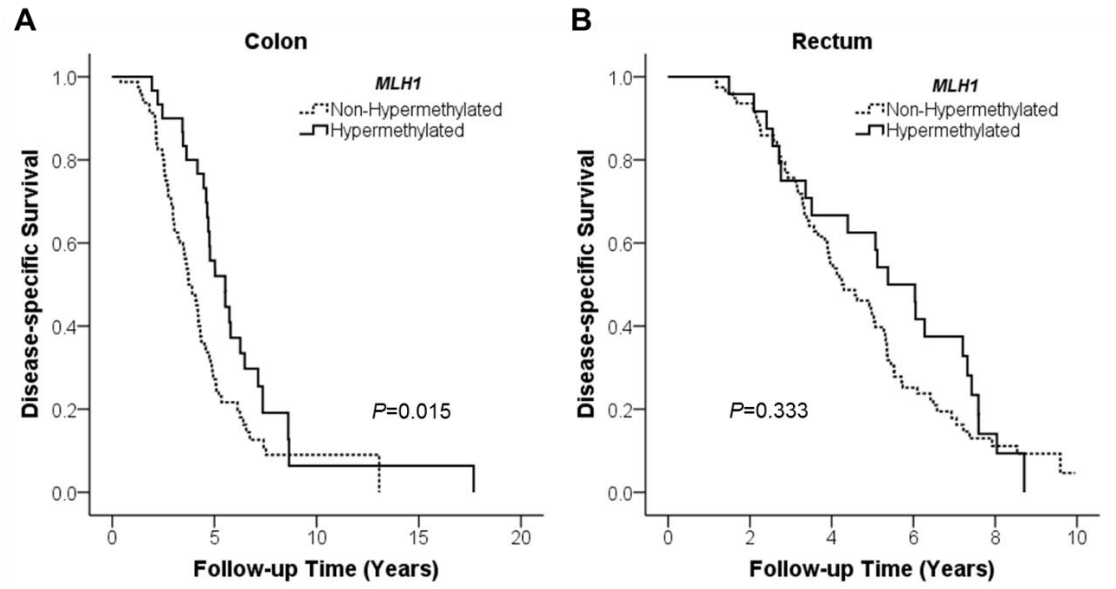

Figure 18 – Kaplan-Meier analysis with the association between DSS and methylation levels in *MLH1* on CRC located in colon (A) and in rectum (B). *P* was assessed by Log-Rank Test. (n=110 for colon and n=104 for rectum)

Discussion

CRC is one of the most common and deadly malignancies worldwide [4]. Although biopsy is needed for histological confirmation, CRC's screening for early-stages can only be accomplished by colonoscopy presently. This is an invasive and expensive procedure, despite being essential in diagnosis of several gastrointestinal diseases, including CRC. Currently, no other accurate diagnostic alternative has been found.

Nevertheless, research efforts has been made in this field and the most promising type of biomarkers for CRC's detection have been shown to be based on aberrant methylation of several cancer–related genes [34]. Indeed, some DNA methylation-based biomarkers were already approved by FDA, including ColoVantage®, Epi proColon® and ColoSure®. While the former two are based on *SEPT9* methylation [62, 63], the latter is based on *Vimentin* methylation [65]. The value of *SEPT9* promoter methylation in CRC's detection has been recently questioned, however, by a recent study that showed only 48.2% sensitivity and 91.5% specificity [64]. As for *Vimentin methylation*, the use of this test needs to be combined with colonoscopy since its sensitivity is variable, ranging from 38-88% [65, 66].

In this context, this dissertation aimed to find and validate a novel panel of biomarkers based on DNA promoter methylation. After literature review, six genes were selected and then validated in CRC's tissue samples. Those included two tumor suppressor genes, *APC* and *RASSF1A*, a gene implicated in mismatch repair, *MLH1*, one that encodes for a normally imprinted fetal growth factor, *IGF2*, the *septin* gene, *SEPT9*, and finally, a DNA repair gene, *MGMT*.

Excepting for *MLH1,* the majority of the selected genes were able to distinguish CRC from CRN samples. From these, *IGF2* methylation levels were slightly lower in CRC than CRN, being in accordance with reported data [79], whereas the remainder showed significantly higher methylation levels on CRC, as it was expected. Considering colon and rectum cancer separately, only *SEPT9*, *MGMT* and *RASSF1A* were able to distinguish CRC from CRN in colon and *SEPT9*, *MGMT* and *APC* in rectum. Therefore, DNA methylation in colon and rectum tissues revealed a slightly different signature, differing only on general cancer genes, *APC* and *RASSF1A*, whereas *SEPT9* and *MGMT* methylation was present in both locations.

Considering the methylation levels of CRC samples between the two locations, colon and rectum, no significant differences were found between them, with the exception of *SEPT9* which had a slightly higher methylation on colon samples. Thus,

location does not seem a relevant factor that may limit the ability of these markers to identify CRC.

MSI-H has been reported to be due to defects in the DNA MMR system [40]. This system includes genes like *MLH1*, *MSH2*, *MSH6* and *PMS2* whose expression was screened by immunohistochemistry. Therefore, assuming that the lack of expression, in any of these genes, represents a defect on this system, one may assume that these defects possibly lead to MSI-H pathway and so, the samples on this situation will be considered as MSI-H, whereas the remaining samples most likely represent MSI-L or MSS cases. Knowing that the two types of instability are mutually exclusive, a tumor fitting into the CIN pathway most likely would be MSS [41, 42], therefore, the group MSI-L/MSS would be highly concentrated on samples falling into CIN pathway. The group MSI-L/MSS will, in a grossly way, represent CIN behavior.

Between the two categories, MSI-H and MSI-L/MSS, only *APC* showed differences in methylation levels (P=0.008). All the other genes did not show differences in methylation levels and so it can be assumed that, in a general way, the selected genes were able to detect CRC both in MSI or CIN pathways. Thus, only remains to know if the same is verified in samples fitting the CIMP pathway. The CIMP characterization is being done in a parallel project.

In this cohort, it was observed that, for rectum only, *SEPT9* had higher methylation levels for stage IV when compared to stage I, II or III (P=0.001) and for both *SEPT9* and *MGMT*, methylation levels were lower for patients that received neoadjuvant treatment (P=0.002 and P=0.012, respectively). In fact, there are several studies reporting and demonstrating that ionizing radiation exposure could affect DNA methylation patterns. The most frequently reported radiation-induced situations are global hypomethylation, loss of methylation, paralleled with a decrease in expression levels of methyltransferases and methyl CpG binding proteins [95, 96]. In colon cancer a trend for radiation-induced hypomethylation had also been reported [97].

Methylation levels observed for *MGMT* and *MLH1* were higher in female both in colon and rectum samples, a finding that might be related with the reported higher *MLH1* promoter methylation in female patients already reported in a meta-analysis study [98].

For *MLH1* or any of the other genes, no significant differences were found among the age groups, suggesting that the methylation sites analyzed do not endure significant age-related changes in DNA methylation levels. Age-related methylation in colorectal tissues has been extensively, although this phenomenon is not exclusive of colorectal tissues, as other human tissues also display this type of methylation [99, 100]. The possibility of hypermethylation observed in cancer be age-related was raise, but it was concluded that there were two types of methylation in cancer, the age-related

methylation, which precedes neoplastic transformation, and cancer-related methylation, which is totally cancer-related [99]. No correlation between age and selected genes methylation was found in the samples of this cohort, which suggests that the methylation found most certainly corresponds to the cancer-related type.

The gene with the best performance in terms of diagnosis was *SEPT9* followed by *MGMT*. *SEPT9* showed 85.5% sensitivity and 94.0% specificity, which is not very different from the data presented by the trademark assays using *SEPT9* methylation [62, 101]. *MGMT* displayed a better performance, with 77.2% sensitivity and 84.0% specificity, which is superior to that previously reported: 46-53% sensitivity and 74-100% specificity, both in tissue samples [84, 102].

APC showed 35.2% sensitivity and 90.0% specificity which comes close to the 47% sensitivity and 77% specificity [103], and 97% specificity [69] previously described by other authors. Although the results observed for *RASSF1A*, 33.1% sensitivity and 90.0% specificity, differ from those reported for tissue samples (81.0% sensitivity and 51.0% specificity [74]), it should be recalled that different methodologies were used, precluding direct comparisons. The values obtained to *IGF2* were lower than it would be expected, 58.6% sensitivity and 64.0% specificity against 80.0% sensitivity and 90.0% specificity reported previously, although pyrosequencing was used instead of MSP [81]. For *MLH1*, with 13.8% sensitivity and 97.9% specificity, sensitivity was much lower than previously reported (95% [104]) using sequencing of nested MSP products, although specificity was similar to the one previously described, 97.0% [69] (using MSP).

Combining *SEPT9* and *MGMT*, an improvement of sensitivity could be reached, with 93.8% sensitivity, 82.0% specificity and 82.0% NPV. Adding *RASSF1A* to the panel might improve its diagnostic performance through the acquisition of good balance between all parameters: 96.6% sensitivity, 74.0% specificity, 91.5% PPV and 72.5% NPV.

The *SEPT9/MGMT/RASSF1A* panel was able to detect CRC both in colon (proximal and distal) and rectum with a positive detection rate above 90%, more exactly, 95.7% for colon and 98.0% for rectum. Besides, the panel was also able to detect equally tumors at any disease stage, with a positive detection rate of 100% for stage I and II, 94.2% for stage III and 95.9% for stage IV. The panel displayed better positive detection rates than *SEPT9* alone, supporting the theory that a panel has a better performance on CRC diagnosis.

Some of the analyzed genes in this study displayed prognostic value. Indeed, hypermethylation of *SEPT9* (DFS HR=0.031, and DSS HR=0.031) and *MLH1* (DFS HR=0.027, and DSS HR=0.033) were associated with a better prognosis, i.e., increased DFS and DSS. For DSS they were also independent prognostic factors. Considering

colon and rectum separately, none of the genes had prognosis value for DFS, as for DSS, although hypermethylation of *SEPT9* was an independent factor for colon (*P*=0.006) and it was associated with a better prognosis on both locations. Hypermethylation of *MLH1* was an independent factor associated with a better prognosis exclusively for tumors located on colon (*P*=0.014). For rectum, no association with prognosis was found for *MLH1* methylation. *MGMT* did not show any association with CRC prognosis as already reported in a meta-analysis that included fourteen studies [105].

Conclusions and Future Perspectives

Globally, for detection/diagnostic purposes, our results parallel the published data, mostly regarding *SEPT9*, *MGMT* and *RASSF1A* promoter methylation in CRC.

The gene with best diagnostic value was *SEPT9*, similar to some published data, supporting *SEPT9*'s promoter methylation as a promising biomarker, showing high sensitivity for CRC. In addition, these results suggest that *MGMT* is a very good biomarker as well, and when put together with *SEPT9*, the CRC detection sensitivity is highly improved. The addition of *RASSF1A* to the panel contributes for the cancer baseline detection, improving the panel robustness. Thus, a panel combining *SEPT9/MGMT/RASSF1A* might provide superior performance than currently available epigenetic-based biomarkers for CRC. This shows that it is possible to improve CRC detection efficacy by combining some well-known biomarkers already reported individually.

Methylation levels in tissues following MSI and CIN pathways were not significantly different in the two major biomarkers of the panel (*SEPT9* and *MGMT*) which suggest that CRC detection by this panel will not be molecular pathway's dependent, at least for these two paths. Remains to know if the same is verified for CIMP.

This dissertation demonstrated that promoter methylation of *SEPT9/MGMT/RASSF1A* might constitute a useful panel for CRC identification. However, to be used in a non-invasive setting, it must be tested in easily obtainable biofluids, especially blood. However, the specificity of this approach will be challenged by the possibility that other cancers, especially the most common (e.g., gastric, pulmonary, pancreatic) might also share similar DNA methylation signatures. Thus, future studies should also include as "controls" not only healthy individuals, but also carriers of other cancer types.

Appendix 1: Genes Selection - Literature review

In order to find genes with relevance in this matter, a literature review was made. For start, it was compiled a list of genes reported to be methylated in CRC, only the ones with significant cohorts. The sample type used for methylation assessment, sensitivity and specificity reported in each article was annotated in the same list. All this information was taken into account to the selection of potential genes (Table 19).

However, displaying methylation on CRC cases was not enough, it was needed that the selected genes were specific for CRC, and so, have a very low frequency in other malignancies. For this reason, for some of the genes, it was search the frequency observed in other types of cancer. On Table 20 is shown the results of this search, displaying information about the type of cancer in which methylation is also present, its frequency, the specimen type and the sample size.

The aim of this study is not only found good biomarkers individually but also to find a panel with a great performance. Therefore, a search for panels used in methylation analysis was made in order to assess the ones that had already been reported. Information about the sensitivity, specificity, specimen type and method used was collected when available, Table 21.

Considering all this search, some genes demonstrated potential usefulness, not only to detect globally cancer, such as *APC* and *RASSF1A*, but also to detect CRC, such as *MLH1*, *MGMT*, *SEPT9* and *IGF2*.

The selection of the gene location to perform this study was a very careful one. For some of the genes were used the same primers as described in literature, for others, the primers were design specifically for a very exact location.

In order to study the promoter methylation of *APC*, primers in the same gene location that previous studies [68, 103, 106] were design. The primers used for RASSF1A were designed in the same gene region reported by other authors [74, 103, 107].

Table 19 – List of genes reported to be methylated in CRC. *IGF2* is described to be hypomethylated. * hypomethylation ≤35%.

Gene	Sensitivity (%)	Specificity (%)	Specimen Type	References
DCLK1	82	100	Tissue	[108]
SEPT9	69	86	Plasma	[109]
TMEFF2	65	69	Plasma	[109]
	-	100	Plasma	[110]
SFRP1	52	92	Stool	[111]
SFRP2	93	100	Tissue	[102]
	94	96	Stool	[102]
	57	90	Stool	[112]
MGMT	53	100	Tissue	[102]
	48	100	Stool	[102]
	-	96	Tissue	[69]
	-	100	Plasma	[69]
	46	74	Tissue	[84]
RASSF2A	60	100	Tissue	[69]
	-	100	Plasma	[69]
	84	98	Stool	[113]
	73	88	Tissue	[114]
	70	100	Tissue	[115]
HLTF	-	98	Tissue	[69]
	41	93	Plasma	[69]
	52	82	Tissue	[103]
MLH1	95	-	Tissue	[104]
	-	97	Tissue	[69]
	-	100	Plasma	[69]
CDKN2A	73	-	Tissue	[116]
	71	-	Plasma	[116]
	-	97	Tissue	[69]
	-	100	Plasma	[69]
NGFR	51	84	Plasma	[109]
HPP1	73	100	Tissue	[102]
	71	100	Stool	[102]
HIC1	42	98	Stool	[117]
UNC5C	68	-	Tissue	[118]
DCC	56	-	Tissue	[118]
RUNX3	75.6	76.3	Tissue	[119]
	34	-	Tissue	[120]
Wif-1	74	98	Tissue	[69]
	37	91	Plasma	
GATA4	70	94	Tissue	[103]
	51	93	Stool	[103]
	43	95	Stool	[112]
GATA5	79	87	Tissue	[103]
	84	83	Stool	[112]
OSMR	32	-	Tissue	[118]
DFNA5	34	-	Tissue	[66]
RASSF1A	81	51	Tissue	[74]
*IGF2**	80	90	Tissue	[81]
VIM	41	85	Stool	[112]
APC	-	97	Tissue	[69]
	-	100	Plasma	[69]
	47	77	Tissue	[103]

Table 20 – Methylation frequency in other malignancies for some of the previous genes.

Gene	Cancer type	Freq. (%)	Specimen type	Sample size	References
RASSF2A	Gastric carcinoma	19	Fecal	n=21	[113]
IGF2	Colorectal cancer	80	Tissue	n=42	[81]
	Breast cancer	33	Tissue	n=22	
GATA4	Lung cancer	67	Tissue	n=63	[121]
	Esophageal adenocarcinoma	71	Tissue	n=44	[122]
GATA5	Lung cancer	41	Tissue	n=63	[121]
	Esophageal adenocarcinoma	55	Tissue	n=44	[122]
Wif-1	Esophageal cancer	80	Tissue	n=20	[123]
	Gastric cancer	74.2	Tissue	n=31	
	Colorectal cancer	82	Tissue	n=50	
	Pancreatic cancer	75	Tissue	n=8	
	Lung cancer	83	Tissue	n=18	[124]
RUNX3	Breast carcinoma	47	Plasma	n=19	[125]
	Non-small cell lung cancer	55		n=20	
	Gastric carcinoma	100		n=4	
	Pancreatic carcinoma	100		n=2	
	Colorectal carcinoma	65		n=17	
	Liver carcinoma	88		n=8	
	Non-small cell lung cancer	25	Tissue	n=101	[126]
	Breast cancer	53	Tissue	n=20	[127]
	Gastric carcinoma	71	Tissue	n=80	[128]
RASSF1A	Breast carcinoma	42	Plasma	n=19	[125]
	Non-small cell lung	30		n=20	
	Gastric carcinoma	25		n=4	
	Pancreatic carcinoma	50		n=2	
	Colorectal carcinoma	24		n=17	
	Liver carcinoma	50		n=8	
	Lung cancer	88	-	-	Reviewed [129]
	Breast cancer	95	-	-	
	Prostate cancer	99	-	-	
	Parathyroid tumor	98	Tissue	n=55	[130]
	Non-small cell lung cancer	42	Tissue	n=101	[126]
APC	Breast cancer	36	Tissue	n=50	[70]
	Pancreatic acinar cell carcinoma	56	Tissue	n=43	[131]
	Breast cancer	45	Tissue	n=84	[132]
		31	Serum	n=84	
	Parathyroid tumor	71	Tissue	n=55	[130]
	Esophageal adenocarcinoma	78	Tissue	n=50	[133]
	Cardiac adenocarcinoma	32	Tissue	n=50	
	Gastric adenocarcinoma	84	Tissue	n=50	
MLH1	Endometrial carcinoma	45	Tissue	n=29	[77]
MGMT	Oral squamous cell carcinoma	41	Tissue	n=99	[89]
	Non-small cell lung cancer	14	Tissue	n=101	[126]
	Gliomas	64.8	Tissue	n=247	[134]

Table 21 – Panels of genes based on DNA methylation reported in a few studies. COBRA: Combined Bisulfite Restriction Analysis. Hi-SA: High-sensitivity assay for bisulfite DNA. qMSP: Quantitative Methylation-specific PCR. QuARTS: is a Real-time target and signal amplification technology.

Panel of markers	Sensitivity (%)	Specificity (%)	Specimen type	Method	Ref.
TMEFF2; NGFR; SEPT9	-	-	Plasma	Microarray; qPCR	[109]
APC; MGMT; RASSF2A; Wif-1	86.5	92.1	Plasma	MSP	[69]
SFRP2; RASSF2	75.0	89.4	Stool	COBRA, Hi-SA	[113]
BMP3; NDRG4; TFPI2; Vimentin	85	89	Stool	QuARTS	[135]
HPP1; HLTF; MLH1	-	-	Plasma	MSP	[136]
UNC5C; DCC	82	-	Tissue	qMSP	[118]
SFRP2; HPP1; MGMT	93.7	77.1	Stool	MSP	[102]

The promoter of *MLH1* is divided into four regions, A, B, C and D, and it was shown that the absence of *MLH1* expression was more correlated with the methylation within C region. Methylation of A and B regions were not specific for loss of expression and in D region was also a good indicator of *MLH1* silencing, despite being less specific than C region [75, 137]. The absence of *MLH1* expression is an indicator of MSI, for this study we already have this information by immunohistochemistry. For this motive and with the intent of have more information about the methylation on promoter region, the primers used were the same described by Widschwendter, Siegmund [138].

IGF2 LOI is correlated with hypomethylation at a differentially methylated region (DMR)-0 [79]. With the intention of studying the methylation status of DMR-0 (Figure 19), primers were design within this region including the only three CpGs [78, 79] on forward primer.

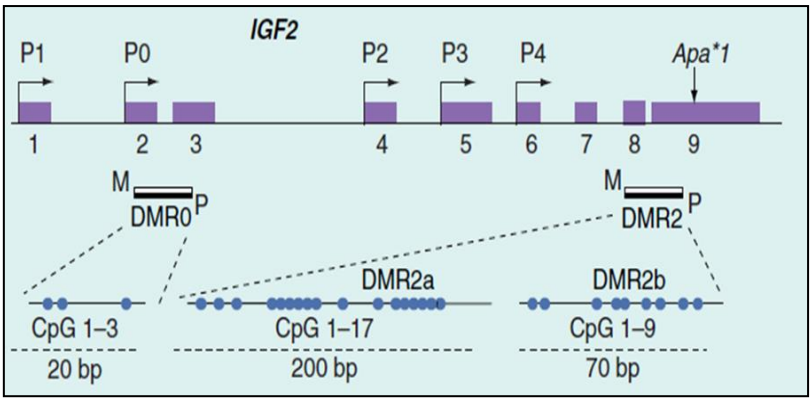

Figure 19 – CpGs present in each DMR of *IGF2*. P1-P4 represent the four promotors of the gene (Adapted from [79])

Within *SEPT9* gene were found four CpG islands, one extragenic and three intragenic, they were designated as CGI1, CGI2, CGI3 and CGI4 (Figure 20). The only one described as differentially methylated in CRC is CGI3. A study by Wasserkort, Kalmar [82] assessing which location within CGI3 is more specific for neoplasia, tested three sets of primers in this region. The primers for the amplicon located centrally in CGI3 (Figure 20) have shown the largest methylation differences between normal and neoplastic samples [82]. This is the target region used by diagnostic tests currently available for clinical application [82]. For this reason, primers within the ones used by this author were designed and used in the assessment of *SEPT9* promoter methylation levels.

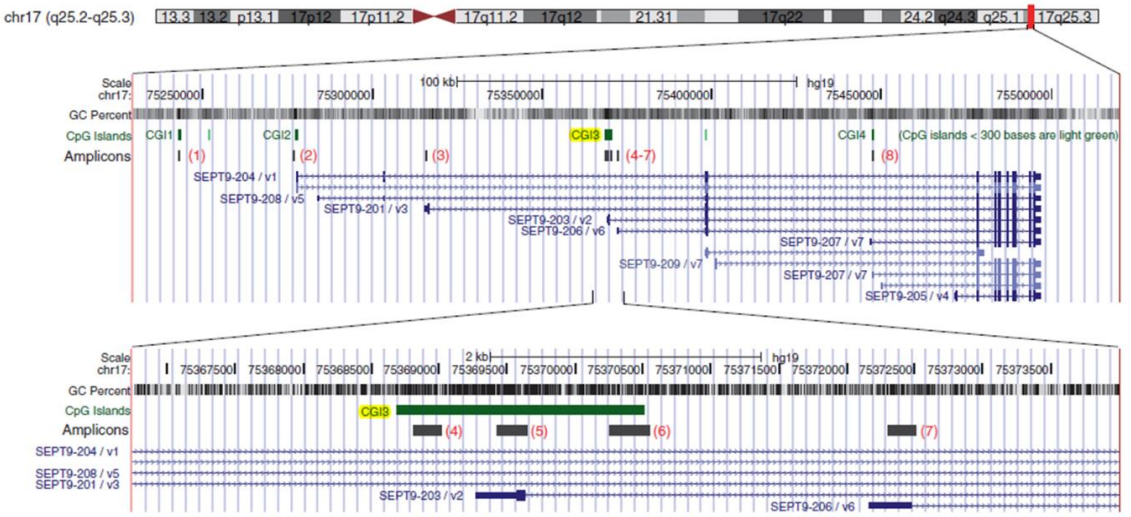

Figure 20 – Genomic organization of SEPT9. Location of each CGI and the detail of CGI3 with the location of the three amplicons. (5) is the amplicon which the methylation was more specific for CRC detection (Adapted from [82])

For *MGMT*, the primers used were the ones described by Huang, Li [102] and Ramalho-Carvalho, Pires [86] which are located in the same region as primers described by other authors [46, 138]. The primers used were located in the promoter region in which methylation is more related with loss of *MGMT* expression [86].

Appendix 2: Clinical data base construction

In order to do a complete evaluation of diagnostic and prognostic value of each one of the selected biomarkers, clinical information of all of the cohort's patients was collected and displayed in a data base. This information was mostly available on patients' clinical charts on virtual or paper processes.

Some basic information about each patient was first collected. On this data was included: gender, date of birth, age at diagnosis, tumor location (Rectum, Distal Colon and Proximal colon), confirmation of diagnosis (Adenocarcinoma), date of surgery, age at surgery, if the patient had neoplastic history, personal or familiar, and when applicable, which type of tumor is included in this history. In the location feature, proximal colon included tumors on cecum, ascending colon, hepatic flexure and transverse colon, while distal colon included tumors on splenic flexure, descending colon and sigmoid colon.

Next, information about the tumor and specifically about the tissue used in this study was collected too. In these were included: histologic grade, microscopic tumor extension, if the surgical margins were tumor-free (proximal, distal and radial), number of total lymph nodes removed during surgery and how many were metastasized, if had metastasis *ad initio* and where (Liver, Lung or in other location), TNM classification (pathological T and N; M were pathological or clinical depending on the case), stage and if the patient did neoadjuvant treatment and which regimen. This data was collected take into account the information displayed on "staging" chapter (Introduction). Considering the information obtain about the Lymph nodes, the ratio of metastasized lymph nodes was determined. Follow-up information was also collected: current treatment (when the patient was still alive), if did adjuvant treatment and which regimen, if were used biologic agents in adjuvant therapy, if metastasized after surgery, which organ and if had local recurrence. Data regarding *KRAS*, *BRAF* and *PIK3CA* mutations, when available, was also collected. For the *PIK3CA* mutation, nothing was found.

For prognosis analysis the information collected was for survival analysis: date of diagnosis, vital status and date of last follow-up or death depending on vital status; and for disease-free survival: date of last curative treatment, if the patient relapse, date of recurrence or date of follow-up/death if the patient didn't relapse and is alive or die disease-free.

Some of these features were reviewed by a pathologist to ensure its accuracy.

Appendix 3: Results I

Table 22 – Analyses of clinical and pathological variables from Colon for each one of the genes in study (*P* values were obtained by Mann-Whitney Test or Kruskal-Wallis Test).

n=110 (**Colon**)			***SEPT9* methylation**		***MGMT* methylation**		***RASSF1A* methylation**		***APC* methylation**		***IGF2* methylation**		***MLH1* methylation**	
Variables		n	Median	*P*	Median	*P*	Median	*P*	Median	*P*	Median	*P*	Median	*P*
Sex	Female	42	236.448	0.204	48.121	**0.048**	0.000	0.516	13.234	0.917	63.036	0.214	2.660	**0.007**
	Male	68	153.980		20.967		0.000		15.577		57.953		0.000	
Age Group	18-56	35	166.062	0.757	18.274	0.311	0.000	0.355	15.490	0.241	58.208	0.913	0.000	0.995
	57-65	40	179.207		35.050		0.000		18.777		60.777		1.249	
	66-85	35	217.652		22.159		0.000		9.338		60.826		1.419	
Mucinous histology	No	98	212.920	0.104	26.536	0.283	0.000	0.991	14.801	0.946	58.936	0.079	1.342	0.127
	Yes	12	123.236		13.176		0.000		14.409		74.254		0.000	
Stage	I and II	24	225.064	0.348	41.618	0.282	0.000	0.257	15.685	0.716	52.354	0.456	1.959	0.448
	III	20	193.869		29.731		0.000		11.709		62.703		1.794	
	IV	64	179.207		20.967		0.000		13.666		63.017		0.793	
Primary Tumor (pT)	T1 and T2	6	397.610	**0.018**	33.907	0.974	0.007	0.420	12.281	0.894	56.790	0.160	3.189	0.052
	T3	89	179.443		25.038		0.000		14.861		59.823		1.332	
	T4	13	111.014		47.448		0.000		14.741		90.308		0.000	
	Tx	2	906.720		112.569		8.981		208.568		29.109		7.980	
Regional Lymph Nodes (pN)	N0	39	197.867	**0.000**	36.510	0.554	0.000	0.932	15.490	0.682	48.382	**0.046**	1.328	0.836
	N1	33	251.779		20.922		0.000		9.338		64.305		1.503	
	N2	36	102.014		18.592		0.000		18.229		63.488		0.000	
	Nx	2	906.720		112.569		8.981		208.568		29.109		7.980	
Distant Metastasis (M)	M0	44	224.365	**0.011**	38.285	0.370	0.000	0.407	15.116	0.224	59.743	0.056	1.880	0.225
	M1	51	210.815		18.312		0.000		10.770		58.208		1.225	
	M2	13	51.714		44.147		0.000		35.887		153.440		0.000	
	Mx	3	906.720		112.569		8.981		208.568		29.109		7.980	
Neoadjuvant Treatment	No	94	197.120	0.514	25.619	0.482	0.000	0.165	16.210	0.084	59.824	0.536	1.049	0.217
	Yes	16	169.485		28.136		0.000		10.320		61.958		2.142	
KRAS Mutation: resistance	NA	66	133.552	0.064	18.861	0.022	0.000	0.067	15.879	0.950	63.885	0.688	0.000	0.564
	Yes	44	228.674		42.814		0.000		11.982		59.016		1.498	

Table 23 - Analyses of clinical and pathological variables from Rectum for each one of the genes in study (*P* values were obtained by Mann-Whitney Test or Kruskal-Wallis Test).

n=104 (**Rectum**)			***SEPT9* methylation**		***MGMT* methylation**		***RASSF1A* methylation**		***APC* methylation**		***IGF2* methylation**		***MLH1* methylation**	
Variables		n	Median	*P*	Median	*P*	Median	*P*	Median	*P*	Median	*P*	Median	*P*
Sex	Female	32	145.921	0.781	54.394	**0.049**	0.000	0.199	20.028	0.443	75.824	0.510	2.248	**0.010**
	Male	72	107.688		21.499		0.000		14.251		65.966		0.000	
Age Group	18-56	34	118.231	0.451	20.158	0.452	0.000	0.270	17.542	0.235	86.093	0.059	0.936	0.680
	57-65	41	145.092		26.239		0.000		22.341		67.232		0.864	
	66-85	29	93.747		31.629		0.000		9.535		54.145		0.803	
Mucinous histology	No	100	127.550	0.288	27.133	0.112		0.296	15.591	0.125	67.232	0.296	0.803	0.275
	Yes	**4**	79.959		8.128		0.000		7.774		77.095		17.957	
Stage	I and II	28	94.972	**0.001**	33.702	0.183	0.000	0.646	15.214	0.716	68.874	0.513	1.702	0.535
	III	32	53.125		24.307		0.000		14.948		72.671		1.814	
	IV	44	179.276		28.626		0.000		15.591		61.247		0.000	
Primary Tumor (pT)	T1 and T2	16	95.599	0.820	32.041	0.714	0.000	0.362	14.832	0.525	86.018	0.668	1.260	0.462
	T3	84	123.484		24.412		0.000		14.989		64.811		0.000	
	T4	4	267.495		105.671		0.269		26.666		71.659		5.188	
	Tx	0	-		-		-		-		-		-	
Regional Lymph Nodes (pN)	N0	38	119.159	0.788	52.717	0.266	0.000	0.597	19.050	0.969	65.388	0.913	0.958	0.758
	N1	32	131.317		17.908		0.000		15.591		68.364		0.000	
	N2	30	120.858		24.879		0.000		14.238		69.570		1.396	
	Nx	4	303.043		20.790		-		9.156		66.709		2.269	
Distant Metastasis (M)	M0	60	75.598	**0.000**	25.264	0.151	0.000	0.766	15.214	0.633	70.148	0.270	1.704	0.554
	M1	37	200.538		28.980		0.000		16.689		62.746		0.000	
	M2	7	123.484		23.647		0.000		10.794		24.786		0.000	
	Mx	0	-		-		-		-		-			
Neoadjuvant Treatment	No	51	163.200	**0.002**	31.629	**0.012**	0.000	0.802	14.989	0.403	65.966	0.974	0.980	0.547
	Yes	53	91.597		16.099		0.000		15.514		68.020		0.402	
KRAS Mutation: resistence	NA	66	131.317	0.242	22.986	0.110	0.000	0.552	15.438	0.949	66.186	0.656	0.980	0.614
	Yes	37	101.752		49.714		0.000		14.989		69.779		0.000	

Appendix 4: Results II

Table 24 – Cox regression assessing the potential of methylation in prediction of Disease-free survival (DFS) and disease-specific survival (DSS). * Non-hypomethylated used as indicator, all the other genes had non-hypermethylated as indicator, n=214.

Univariable analysis		DFS		DSS	
Gene	**Methylation**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SEPT9	Hyper	0.651 (0416-1.019)	0.060	0.721 (0.516-1.006)	**0.054**
MGMT	Hyper	0.709 (0.424-1.185)	0.189	0.915 (0.656-1.274)	0.598
RASSF1A	Hyper	0.884 (0.575-1.358)	0.573	1.011 (0.727-1.407)	0.946
IGF2	Hypo*	0.968 (0.592-1.586)	0.899	0.928 (0.655-1.314)	0.673
APC	Hyper	0.943 (0.593-1.500)	0.805	0.835 (0.592-1.179)	0.306
MLH1	Hyper	0.669 (0.434-1.033)	0.070	0.665 (0.476-0.930)	**0.017**

Table 25 – Cox regression assessing the potential of clinical features in prediction of Disease-free survival (DFS) and disease-specific survival (DSS), n=214.

Univariable analysis			DFS		DSS	
		n	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Stage	I and II	52	1 (referent)	0.304	1 (referent)	**0.000**
	III	51	1.392 (0.896-2.162)	0.141	0.990 (0.651-1.506)	0.962
	IV	108	1.027 (0.591-1.785)	0.925	1.900 (1.323-2.728)	**0.001**
Primary Tumor (pT)	T1 and T2	22	1 (referent)	**0.013**	1 (referent)	**0.000**
	T3	172	1.437 (0.798-2.587)	0.227	1.050 (0.664-1.661)	0.835
	T4a and T4b	17	5.826 (1.800-18.860)	**0.003**	3.288 (1.711-6.319)	**0.000**
Regional Lymph Nodes (pN)	N0	77	1 (referent)	0.261	1 (referent)	**0.040**
	N1 a, b, c	64	1.066 (0.673-1.688)	0.785	1.012 (0.703-1.457)	0.948
	N2 a,b	66	1.668 (0.989-2.813)	0.055	1.559 (1.100-2.209)	**0.013**
Distant Metastasis (M)	M0	103	1 (referent)	0.898	1 (referent)	**0.000**
	M1a	88	1.129 (0.675-1.889)	0.645	1.884 (1.382-2.568)	**0.000**
	M1b	20	-	-	2.049 (1.224-3.431)	**0.006**
Gender	Male	139	1 (referent)	-	1 (referent)	-
	Female	74	0.686 (0.455-1.034)	0.072	1.023 (0.755-1.387)	0.883
Neoadjuvant Treatment	No	145	1 (referent)	-	1 (referent)	-
	Yes	68	1.224 (0.797-1.881)	0.356	1.365 (1.007-1.851)	**0.045**
Age group	18-57	76	1 (referent)	0.080	1 (referent)	**0.001**
	58-65	73	1.709 (1.064-2.746)	**0.027**	1.635 (1.146-2.332)	**0.007**
	66-85	64	1.179 (0.717-1.938)	0.517	1.994 (1.377-2.888)	**0.000**
Colon	Rectum	103	1 (referent)	0.085	1 (referent)	0.161
	Distal	73	0.647 (0.408-1.024)	0.063	0.979 (0.705-1.360)	0.901
	Proximal	37	1.183 (0.685-2.042)	0.547	1.429 (0.961-2.125)	0.078

Table 26 – Cox regression assessing the potential of methylation in prediction of disease-free survival (DFS) both in colon and rectum tumors. * Non-hypomethylated used as indicator, all the other genes had non-hypermethylated as indicator, n=110 for colon and n=104 for rectum.

Univariable analysis		Colon		Rectum	
Gene	**Methylation**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SEPT9	Hyper	0.600 (0.317-1.137)	0.117	0.653 (0.328-1.301)	0.225
MGMT	Hyper	0.748 (0.366-1.529)	0.426	0.647 (0.303-1.381)	0.260
RASSF1A	Hyper	1.016 (0.553-1.863)	0.960	0.683 (0.365-1.276)	0.232
IGF2	Hypo*	1.043 (0.515-2.112)	0.908	0.939 (0.467-1.885)	0.858
APC	Hyper	0.854 (0.441-1.655)	0.640	1.126 (0.574-2.208)	0.730
MLH1	Hyper	0.685 (0.368-1.276)	0.233	0.608 (0.328-1.129)	0.115

Table 27 – Cox regression assessing the potential of clinical features in prediction of disease-free survival (DFS) both in colon and rectum tumors, n=110 for colon and n=104 for rectum.

Univariable analysis		Colon			Rectum		
		n	**HR (95% CI)**	***P***	**n**	**HR (95% CI)**	***P***
Stage	I and II	24	1 (referent)	0.334	28	1 (referent)	0.939
	III	20	1.670 (0.847-3.293)	0.139	32	1.078 (0.598-1.943)	0.802
	IV	64	1.193 (0.572-2.486)	0.638	44	0.939 (0.397-2.222)	0.887
Primary Tumor (pT)	T1 and T2	6	1 (referent)	**0.025**	16	1 (referent)	0.223
	T3	89	2.124 (0.829-5.444)	0.117	84	0.868 (0.397-1.896)	0.722
	T4a and T4b	13	11.44(1.947-67.229)	**0.007**	4	3.201 (0.638-16.066)	0.158
Regional Lymph Nodes (pN)	N0	39	1 (referent)	0.344	38	1 (referent)	0.579
	N1 a, b, c	33	1.194 (0.606-2.356)	0.608	32	0.898 (0.475-1.698)	0.740
	N2 a,b	36	1.751 (0.825-3.715)	0.144	30	1.513 (0.725-3.155)	0.270
Distant Metastasis (M)	M0	44	1 (referent)	0.565	60	1 (referent)	0.802
	M1a	51	1.456 (0.730-2.904)	0.286	37	0.903 (0.405-2.010)	0.802
	M1b	13	-	-	7	-	-
Gender	Male	68	1 (referent)	-	72	1 (referent)	-
	Female	42	1.085 (0.603-1.954)	0.785	32	0.296 (0.149-0.589)	**0.001**
Neoadjuvant Treatment	No	94	1 (referent)	-	51	1 (referent)	-
	Yes	16	1.633 (0.580-4.601)	0.353	53	1.030 (0.596-1.778)	0.916
Age group	18-57	38	1 (referent)	0.311	39	1 (referent)	0.343
	58-65	37	1.718 (0.842-3.503)	0.137	36	1.601 (0.848-3.021)	0.147
	66-85	35	1.151 (0.566-2.340)	0.697	29	1.172 (0.574-2.392)	0.664
Colon	Proximal	37	1 (referent)	0.062	-	-	-
	Distal	73	0.563 (0.307-1.030)	0.062	-	-	-

Table 28 – Cox regression assessing the potential of methylation in prediction of disease-specific survival (DSS) both in colon and rectum tumors. * Non-hypomethylated used as indicator, all the other genes had non-hypermethylated as indicator, n=110 for colon and n=104 for rectum.

Univariable analysis		Colon		Rectum	
Gene	**Methylation**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SEPT9	Hyper	0.624 (0.394-0.988)	**0.044**	0.805 (0.492-1.316)	0.386
MGMT	Hyper	0.871 (0.543-1.398)	0.568	0.976 (0.611-1.556)	0.917
RASSF1A	Hyper	0.861 (0.546-1.357)	0.518	1.170 (0.722-1.896)	0.524
IGF2	Hypo*	1.115 (0.683-1.821)	0.663	0.778 (0.473-1.279)	0.322
APC	Hyper	0.704 (0.436-1.138)	0.152	0.992 (0.602-1.632)	0.973
MLH1	Hyper	0.563 (0.352-0.898)	**0.016**	0.787 (0.485-1.279)	0.334

Table 29 – Cox regression assessing the potential of clinical features in prediction of disease-specific survival (DSS) both in colon and rectum tumors, n=110 for colon and n=104 for rectum.

Univariable analysis		Colon			Rectum		
		n	**HR (95% CI)**	***P***	**n**	**HR (95% CI)**	***P***
Stage	I and II	24	1 (referent)	**0.008**	28	1 (referent)	**0.005**
	III	20	0.800 (0.411-1.556)	0.511	32	1.171 (0.670-2.046)	0.579
	IV	64	1.758 (1.060-2.917)	**0.029**	44	2.208 (1.289-3.784)	**0.004**
Primary Tumor (pT)	T1 and T2	6	1 (referent)	**0.000**	16	1 (referent)	0.201
	T3	89	1.814 (0.727-4.526)	0.201	84	0.790 (0.457-1.364)	0.397
	T4a and T4b	13	6.858 (2.336-20.128)	**0.000**	4	1.880 (0.623-5.673)	0.262
Regional Lymph Nodes (pN)	N0	39	1 (referent)	0.287	38	1 (referent)	0.125
	N1 a, b, c	33	1.043 (0.623-1.749)	0.872	32	1.045 (0.621-1.759)	0.868
	N2 a,b	36	1.557 (0.957-2.534)	0.075	30	1.612 (0.972-2.674)	0.064
Distant Metastasis (M)	M0	44	1 (referent)	**0.024**	60	1 (referent)	**0.003**
	M1a	51	1.957 (1.252-3.060)	**0.003**	37	1.923 (1.226-3.017)	**0.004**
	M1b	13	1.876 (0.966-3.643)	0.063	7	2.967 (1.241-7.092)	**0.014**
Gender	Male	68	1 (referent)	-	72	1 (referent)	-
	Female	42	1.019 (0.667-1.557)	0.931	32	0.980 (0.633-1.519)	0.929
Neoadjuvant Treatment	No	94	1 (referent)	-	51	1 (referent)	-
	Yes	16	2.065 (1.193-3.576)	**0.010**	53	1.307 (0.865-1.975)	0.203
Age group	18-57	38	1 (referent)	**0.021**	39	1 (referent)	0.056
	58-65	37	1.696 (1.007-2.856)	**0.047**	36	1.545 (0.946-2.524)	0.082
	66-85	35	2.061 (1.224-3.469)	**0.007**	29	1.866 (1.096-3.177)	**0.022**
Colon	Proximal	37	1 (referent)	**-**	-	-	-
	Distal	73	0.666 (0.434-1.023)	0.064	-	-	-

References

1. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
2. Fearon, E.R., *Genetic and Epigenetic Alterations in Cancer*, in *Abeloff's Clinical Oncology*. 2014, Elsevier: Philadelphia. p. 188-203.
3. Fitzmaurice, C., et al., *The Global Burden of Cancer 2013.* JAMA oncology, 2015. **1**(4): p. 505-27.
4. Ferlay, J., et al., *Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012.* International Journal of Cancer, 2015. **136**(5): p. E359-E386.
5. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012.* European Journal of Cancer, 2013. **51**(9): p. 1201-1202.
6. Chu, K.M., *1 - Epidemiology and Risk Factors of Colorectal Cancer A2 - Gearhart, Susan L*, in *Early Diagnosis and Treatment of Cancer Series: Colorectal Cancer*, N. Ahuja, Editor. 2011, W.B. Saunders: Saint Louis. p. 1-11.
7. Edge, S., et al., *AJCC Cancer Staging Manual*. 7 ed, ed. S.-V.N. York. 2010, New York: Springer-Verlag XV, 648.
8. Schaeybroeck, S.V., et al., *Colorectal cancer*, in *Abeloff's Clinical Oncology*. 2014, Elsevier: Philadelphia. p. 1278-1335.
9. Carethers, J.M. and B.H. Jung, *Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer.* Gastroenterology, 2015. **149**(5): p. 1177-1190.e3.
10. Giardiello, F.M., *3 - Hereditary Colorectal Cancer and Polyp Syndromes A2 - Gearhart, Susan L*, in *Early Diagnosis and Treatment of Cancer Series: Colorectal Cancer*, N. Ahuja, Editor. 2011, W.B. Saunders: Saint Louis. p. 21-30.
11. Ferri, F.F., *Colorectal Cancer*, in *Ferri's Clinical Advisor 2016*. 2016, Elsevier: Philadelphia. p. 352-354.
12. Imperiale, T.F. and D.F. Ransohoff, *Risk for colorectal cancer in persons with a family history of adenomatous polyps: a systematic review.* Ann Intern Med, 2012. **156**(10): p. 703-9.
13. Lutgens, M.W., et al., *Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies.* Inflamm Bowel Dis, 2013. **19**(4): p. 789-99.
14. Ferlitsch, M., et al., *Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy.* Jama, 2011. **306**(12): p. 1352-8.
15. Deng, L., et al., *Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis.* Dig Dis Sci, 2012. **57**(6): p. 1576-85.
16. Bouvard, V., et al., *Carcinogenicity of consumption of red and processed meat.* Lancet Oncol, 2015. **16**(16): p. 1599-600.
17. Chan, D.S., et al., *Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies.* PLoS One, 2011. **6**(6): p. e20456.
18. Botteri, E., et al., *Smoking and colorectal cancer: a meta-analysis.* Jama, 2008. **300**(23): p. 2765-78.
19. Cho, E., et al., *Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies.* Ann Intern Med, 2004. **140**(8): p. 603-13.
20. Larsson, S.C. and A. Wolk, *Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies.* Am J Clin Nutr, 2007. **86**(3): p. 556-65.
21. Hamilton, W., et al., *Clinical features of colorectal cancer before diagnosis: a population-based case-control study.* Br J Cancer, 2005. **93**(4): p. 399-405.
22. Labianca, R., et al., *Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi64-72.
23. Cunningham, D., et al., *Colorectal cancer.* Lancet, 2010. **375**(9719): p. 1030-47.

24. Benson, A.B., 3rd, et al., *Rectal Cancer, Version 2.2015.* J Natl Compr Canc Netw, 2015. **13**(6): p. 719-28; quiz 728.
25. Benson, A.B., 3rd, et al., *Colon cancer, version 3.2014.* Journal of the National Comprehensive Cancer Network : JNCCN, 2014. **12**(7): p. 1028-59.
26. Glimelius, B., et al., *Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi81-8.
27. Van Cutsem, E., B. Nordlinger, and A. Cervantes, *Advanced colorectal cancer: ESMO Clinical Practice Guidelines for treatment.* Ann Oncol, 2010. **21 Suppl 5**: p. v93-7.
28. Van Cutsem, E., et al., *Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2014. **25 Suppl 3**: p. iii1-9.
29. Goel, A. and C.R. Boland, *Epigenetics of Colorectal Cancer.* Gastroenterology, 2012. **143**(6): p. 1442-+.
30. Vogelstein, B., et al., *Cancer Genome Landscapes.* Science, 2013. **339**(6127): p. 1546-1558.
31. Estecio, M.R.H. and J.P.J. Issa, *Dissecting DNA hypermethylation in cancer.* Febs Letters, 2011. **585**(13): p. 2078-2086.
32. Lao, V.V. and W.M. Grady, *Epigenetics and colorectal cancer.* Nat Rev Gastroenterol Hepatol, 2011. **8**(12): p. 686-700.
33. Gardinergarden, M. and M. Frommer, *CpG Islands in Vertebrate Genomes.* Journal of Molecular Biology, 1987. **196**(2): p. 261-282.
34. Hashimoto, Y., T.J. Zumwalt, and A. Goel, *DNA methylation patterns as noninvasive biomarkers and targets of epigenetic therapies in colorectal cancer.* Epigenomics, 2016.
35. Stoccoro, A., et al., *Epigenetic effects of nano-sized materials.* Toxicology, 2013. **313**(1): p. 3-14.
36. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
37. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes & Development, 2011. **25**(10): p. 1010-1022.
38. Hinoue, T., et al., *Genome-scale analysis of aberrant DNA methylation in colorectal cancer.* Genome Research, 2012. **22**(2): p. 271-282.
39. Choong, M.K. and G. Tsafnat, *Genetic and epigenetic biomarkers of colorectal cancer.* Clin Gastroenterol Hepatol, 2012. **10**(1): p. 9-15.
40. Ogino, S. and A. Goel, *Molecular classification and correlates in colorectal cancer.* J Mol Diagn, 2008. **10**(1): p. 13-27.
41. Jass, J.R., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features.* Histopathology, 2007. **50**(1): p. 113-130.
42. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer.* Cell, 1996. **87**(2): p. 159-70.
43. Whitehall, V.L.J., et al., *Methylation of O–6-Methylguanine DNA Methyltransferase Characterizes a Subset of Colorectal Cancer with Low-level DNA Microsatellite Instability.* Cancer Research, 2001. **61**(3): p. 827-830.
44. Toyota, M., et al., *Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype.* Proceedings of the National Academy of Sciences, 2000. **97**(2): p. 710-715.
45. Hughes, L.A.E., et al., *The CpG island methylator phenotype in colorectal cancer: Progress and problems.* Biochimica Et Biophysica Acta-Reviews on Cancer, 2012. **1825**(1): p. 77-85.
46. Weisenberger, D.J., et al., *CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.* Nature Genetics, 2006. **38**(7): p. 787-793.

47. Ogino, S., et al., *CpG island methylator phenotype-low (CIMP-Low) in colorectal cancer: Possible associations with male sex and KRAS mutations.* Journal of Molecular Diagnostics, 2006. **8**(5): p. 582-588.
48. Shen, L.L., et al., *Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(47): p. 18654-18659.
49. van Rijnsoever, M., et al., *Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands.* Gut, 2002. **51**(6): p. 797-802.
50. Whitehall, V.L.J., et al., *Morphological and Molecular Heterogeneity within Nonmicrosatellite Instability-High Colorectal Cancer.* Cancer Research, 2002. **62**(21): p. 6011-6014.
51. Samowitz, W.S., et al., *Evaluation of a Large, Population-Based Sample Supports a CpG Island Methylator Phenotype in Colon Cancer.* Gastroenterology, 2005. **129**(3): p. 837-845.
52. Ogino, S., et al., *CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies.* Gut, 2006. **55**(7): p. 1000-1006.
53. Ahn, J.B., et al., *DNA Methylation Predicts Recurrence From Resected Stage III Proximal Colon Cancer.* Cancer, 2011. **117**(9): p. 1847-1854.
54. An, B., et al., *Characteristic methylation profile in CpG island methylator phenotype-negative distal colorectal cancers.* International Journal of Cancer, 2010. **127**(9): p. 2095-2105.
55. Noffsinger, A.E., *Serrated Polyps and Colorectal Cancer: New Pathway to Malignancy.* Annual Review of Pathology: Mechanisms of Disease, 2009. **4**(1): p. 343-364.
56. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-767.
57. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development.* N Engl J Med, 1988. **319**(9): p. 525-32.
58. Bettington, M., et al., *The serrated pathway to colorectal carcinoma: current concepts and challenges.* Histopathology, 2013. **62**(3): p. 367-386.
59. Laiho, P., et al., *Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis.* Oncogene, 2007. **26**(2): p. 312-20.
60. Biomarkers Definitions Working, G., *Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework.* Clinical Pharmacology & Therapeutics, 2001. **69**(3): p. 89-95.
61. Shames, D.S., J.D. Minna, and A.F. Gazdar, *Methods for detecting DNA methylation in tumors: From bench to bedside.* Cancer Letters, 2007. **251**(2): p. 187-198.
62. Warren, J.D., et al., *Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer.* Bmc Medicine, 2011. **9**: p. 9.
63. Tóth, K., et al., *Detection of Methylated SEPT9 in Plasma Is a Reliable Screening Method for Both Left- and Right-Sided Colon Cancers.* PLoS ONE, 2012. **7**(9): p. e46000.
64. Church, T.R., et al., *Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer.* Gut, 2014. **63**(2): p. 317-325.
65. Ned, R.M., S. Melillo, and M. Marrone, *Fecal DNA testing for Colorectal Cancer Screening: the ColoSure test.* PLoS currents, 2011. **3**: p. RRN1220.
66. Gyparaki, M.T., E.K. Basdra, and A.G. Papavassiliou, *DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer.* Journal of Molecular Medicine-Jmm, 2013. **91**(11): p. 1249-1256.
67. Fodde, R., et al., *Mutations in the APC tumour suppressor gene cause chromosomal instability.* Nat Cell Biol, 2001. **3**(4): p. 433-438.

68. Esteller, M., et al., *Analysis of Adenomatous Polyposis Coli promoter hypermethylation in human cancer.* Cancer Research, 2000. **60**(16): p. 4366-4371.
69. Lee, B.B., et al., *Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2009. **15**(19): p. 6185-91.
70. Jin, Z., et al., *Adenomatous polyposis coli (APC) gene promoter hypermethylation in primary breast cancers.* British Journal of Cancer, 2001. **85**(1): p. 69-73.
71. Guo, S., et al., *Quantitative assessment of the diagnostic role of APC promoter methylation in non-small cell lung cancer.* Clinical Epigenetics, 2014. **6**(1): p. 5-5.
72. Chen, Y., et al., *APC gene hypermethylation and prostate cancer: a systematic review and meta-analysis.* European Journal of Human Genetics, 2013. **21**(9): p. 929-935.
73. Grawenda, A.M. and E. O'Neill, *Clinical utility of RASSF1A methylation in human malignancies.* British Journal of Cancer, 2015. **113**(3): p. 372-381.
74. Sakamoto, N., et al., *Frequent hypermethylation of RASSF1A in early flat-type colorectal tumors.* Oncogene, 2004. **23**(55): p. 8900-8907.
75. Capel, E., J.F. Flejou, and R. Hamelin, *Assessment of MLH1 promoter methylation in relation to gene expression requires specific analysis.* Oncogene, 2007. **26**(54): p. 7596-7600.
76. Kane, M.F., et al., *Methylation of the hMLH1 Promoter Correlates with Lack of Expression of hMLH1 in Sporadic Colon Tumors and Mismatch Repair-defective Human Tumor Cell Lines.* Cancer Research, 1997. **57**(5): p. 808-811.
77. Esteller, M., et al., *MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas.* Oncogene, 1998. **17**(18): p. 2413-2417.
78. Baba, Y., et al., *Hypomethylation of the IGF2 DMR in Colorectal Tumors, Detected by Bisulfite Pyrosequencing, Is Associated With Poor Prognosis.* Gastroenterology, 2010. **139**(6): p. 1855-1864.
79. Miroglio, A., et al., *Specific hypomethylated CpGs at the IGF2 locus act as an epigenetic biomarker for familial adenomatous polyposis colorectal cancer.* Epigenomics, 2010. **2**(3): p. 365-375.
80. Cheng, Y.W., et al., *Loss of imprinting and marked gene elevation are 2 forms of aberrant IGF2 expression in colorectal cancer.* International Journal of Cancer, 2010. **127**(3): p. 568-577.
81. Ito, Y., et al., *Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer.* Human Molecular Genetics, 2008. **17**(17): p. 2633-2643.
82. Wasserkort, R., et al., *Aberrant septin 9 DNA methylation in colorectal cancer is restricted to a single CpG island.* Bmc Cancer, 2013. **13**: p. 11.
83. Scott, M., et al., *Multimodality expression profiling shows SEPT9 to be overexpressed in a wide range of human tumours.* Oncogene, 2005. **24**(29): p. 4688-4700.
84. Shen, L.L., et al., *MGMT promoter methylation and field defect in sporadic colorectal cancer.* Journal of the National Cancer Institute, 2005. **97**(18): p. 1330-1338.
85. Ogino, S., et al., *Molecular correlates with MGMT promoter methylation and silencing support CpG island methylator phenotype-low (CIMP-low) in colorectal cancer.* Gut, 2007. **56**(11): p. 1564-1571.
86. Ramalho-Carvalho, J., et al., *Altered Expression of MGMT in High-Grade Gliomas Results from the Combined Effect of Epigenetic and Genetic Aberrations.* Plos One, 2013. **8**(3).
87. Yousuf, A., et al., *MGMT gene silencing by promoter hypermethylation in gastric cancer in a high incidence area.* Cellular Oncology, 2014. **37**(4): p. 245-252.

88. Baumann, S., et al., *The prognostic impact of O-6-Methylguanine-DNA Methyltransferase (MGMT) promotor hypermethylation in esophageal adenocarcinoma.* International Journal of Cancer, 2006. **119**(2): p. 264-268.
89. Viswanathan, M., N. Tsuchida, and G. Shanmugam, *Promoter hypermethylation profile of tumor-associated genes p16, p15, hMLH1, MGMT and E-cadherin in oral squamous cell carcinoma.* International Journal of Cancer, 2003. **105**(1): p. 41-46.
90. Clark, S.J., et al., *DNA methylation: bisulphite modification and analysis.* Nat Protoc, 2006. **1**(5): p. 2353-64.
91. Pitt, A.R. and W. Kolch, *Regulation of Gene Expression*, in *Medical Biochemistry*. 2014, Saunders: London. p. 466-485.
92. Navarro, E., et al., *Real-time PCR detection chemistry.* Clinica Chimica Acta, 2015. **439**: p. 231-250.
93. Arya, M., et al., *Basic principles of real-time quantitative PCR.* Expert Review of Molecular Diagnostics, 2005. **5**(2): p. 209-219.
94. Eads, C.A., et al., *MethyLight: a high-throughput assay to measure DNA methylation.* Nucleic Acids Research, 2000. **28**(8): p. e32-00.
95. Hofstetter, B., et al., *Impact of genomic methylation on radiation sensitivity of colorectal carcinoma.* Int J Radiat Oncol Biol Phys, 2010. **76**(5): p. 1512-9.
96. Kim, J.G., et al., *Epigenetics meets radiation biology as a new approach in cancer treatment.* Int J Mol Sci, 2013. **14**(7): p. 15059-73.
97. Bae, J.H., et al., *Identification of radiation-induced aberrant hypomethylation in colon cancer.* BMC Genomics, 2015. **16**: p. 56.
98. Li, X., et al., *MLH1 Promoter Methylation Frequency in Colorectal Cancer Patients and Related Clinicopathological and Molecular Features.* PLoS ONE, 2013. **8**(3): p. e59064.
99. Issa, J.P., *Aging, DNA methylation and cancer.* Crit Rev Oncol Hematol, 1999. **32**(1): p. 31-43.
100. Jones, M.J., S.J. Goodman, and M.S. Kobor, *DNA methylation and healthy human aging.* Aging Cell, 2015. **14**(6): p. 924-32.
101. Wu, D., et al., *Detection of Colorectal Cancer Using a Simplified SEPT9 Gene Methylation Assay Is a Reliable Method for Opportunistic Screening.* J Mol Diagn, 2016. **18**(4): p. 535-45.
102. Huang, Z.H., et al., *Detection of aberrant methylation in fecal DNA as a molecular screening tool for colorectal cancer and precancerous lesions.* World Journal of Gastroenterology, 2007. **13**(6): p. 950-954.
103. Hellebrekers, D., et al., *GATA4 and GATA5 are Potential Tumor Suppressors and Biomarkers in Colorectal Cancer.* Clinical Cancer Research, 2009. **15**(12): p. 3990-3997.
104. Nakagawa, H., et al., *Efficient molecular screening of Lynch syndrome by specific 3 ' promoter methylation of the MLH1 or BRAF mutation in colorectal cancer with high-frequency microsatellite instability.* Oncology Reports, 2009. **21**(6): p. 1577-1583.
105. Li, Y., et al., *Prognostic value of MGMT methylation in colorectal cancer: a meta-analysis and literature review.* Tumour Biol, 2015. **36**(3): p. 1595-601.
106. Enokida, H., et al., *Multigene methylation analysis for detection and staging of prostate cancer.* Clinical Cancer Research, 2005. **11**(18): p. 6582-6588.
107. van Engeland, M., et al., *K-ras mutations and RASSF1A promoter methylation in colorectal cancer.* Oncogene, 2002. **21**(23): p. 3792-3795.
108. Vedeld, H.M., et al., *The recently suggested intestinal cancer stem cell marker DCLK1 is an epigenetic biomarker for colorectal cancer.* Epigenetics, 2014. **9**(3): p. 346-350.
109. Lofton-Day, C., et al., *DNA methylation biomarkers for blood-based colorectal cancer screening.* Clinical Chemistry, 2008. **54**(2): p. 414-423.

110. Sabbioni, S., et al., *Multigene methylation analysis of gastrointestinal tumors: TPEF emerges as a frequent tumor-specific aberrantly methylated marker that can be detected in peripheral blood.* Molecular diagnosis : a journal devoted to the understanding of human disease through the clinical application of molecular biology, 2003. **7**(3-4): p. 201-7.
111. Salehi, R., et al., *Methylation pattern of SFRP1 promoter in stool sample is a potential marker for early detection of colorectal cancer.* Advanced biomedical research, 2012. **1**: p. 87.
112. Lu, H.N., et al., *DNA methylation analysis of SFRP2, GATA4/5, NDRG4 and VIM for the detection of colorectal cancer in fecal DNA.* Oncology Letters, 2014. **8**(4): p. 1751-1756.
113. Nagasaka, T., et al., *Analysis of Fecal DNA Methylation to Detect Gastrointestinal Neoplasia.* Journal of the National Cancer Institute, 2009. **101**(18): p. 1244-1258.
114. Park, H.W., et al., *Correlation between hypermethylation of the RASSF2A promoter and K-ras/BRAF mutations in microsatellite-stable colorectal cancers.* International Journal of Cancer, 2007. **120**(1): p. 7-12.
115. Hesson, L.B., et al., *CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A is an early event in colon carcinogenesis and correlates inversely with K-ras mutations.* Oncogene, 2005. **24**(24): p. 3987-3994.
116. Nakayama, H., et al., *Molecular detection of p16 promoter methylation in the serum of recurrent colorectal cancer patients.* International Journal of Cancer, 2003. **105**(4): p. 491-493.
117. Lenhard, K., et al., *Analysis of promoter methylation in stool: A novel method for the detection of colorectal cancer.* Clinical Gastroenterology and Hepatology, 2005. **3**(2): p. 142-149.
118. Hibi, K., et al., *Aberrant Methylation of the Netrin-1 Receptor Genes UNC5C and DCC Detected in Advanced Colorectal Cancer.* World Journal of Surgery, 2009. **33**(5): p. 1053-1057.
119. Nishio, M., et al., *RUNX3 Promoter Methylation in Colorectal Cancer: Its Relationship with Microsatellite Instability and its Suitability as a Novel Serum Tumor Marker.* Anticancer Research, 2010. **30**(7): p. 2673-2682.
120. Imamura, Y., et al., *RUNX3 promoter region is specifically methylated in poorly-differentiated colorectal cancer.* Anticancer Research, 2005. **25**(4): p. 2627-2630.
121. Guo, M.Z., et al., *Hypermethylation of the GATA genes in lung cancer.* Clinical Cancer Research, 2004. **10**(23): p. 7917-7924.
122. Guo, M.Z., et al., *Hypermethylation of the GATA gene family in esophageal cancer.* International Journal of Cancer, 2006. **119**(9): p. 2078-2083.
123. Taniguchi, H., et al., *Frequent epigenetic inactivation of Wnt inhibitory factor-1 in human gastrointestinal cancers.* Oncogene, 2005. **24**(53): p. 7946-7952.
124. Mazieres, J., et al., *Wnt inhibitory factor-1 is silenced by promoter hypermethylation in human lung cancer.* Cancer Research, 2004. **64**(14): p. 4717-4720.
125. Tan, S.H., et al., *Detection of promoter hypermethylation in serum samples of cancer patients by methylation-specific polymerase chain reaction for tumour suppressor genes including RUNX3.* Oncology Reports, 2007. **18**(5): p. 1225-1230.
126. Yanagawa, N., et al., *Promoter hypermethylation of RASSF1A and RUNX3 genes as an independent prognostic prediction marker in surgically resected non-small cell lung cancers.* Lung Cancer, 2007. **58**(1): p. 131-138.
127. Hwang, K.T., et al., *Downregulation of the RUNX3 gene by promoter hypermethylation and hemizygous deletion in breast cancer.* Journal of Korean Medical Science, 2007. **22**: p. S24-S31.
128. Oshimo, Y., et al., *Frequent loss of RUNX3 expression by promoter hypermethylation in gastric carcinoma.* Pathobiology, 2004. **71**(3): p. 137-143.

129. Donninger, H., M.D. Vos, and G.J. Clark, *The RASSF1A tumor suppressor.* Journal of Cell Science, 2007. **120**(18): p. 3163-3172.
130. Juhlin, C.C., et al., *Frequent Promoter Hypermethylation of the APC and RASSF1A Tumour Suppressors in Parathyroid Tumours.* Plos One, 2010. **5**(3).
131. Furlan, D., et al., *APC alterations are frequently involved in the pathogenesis of acinar cell carcinoma of the pancreas, mainly through gene loss and promoter hypermethylation.* Virchows Archiv, 2014. **464**(5): p. 553-564.
132. Zhang, J.-J., et al., *Detection and significance of APC gene promoter hypermethylation in serum of breast cancer patients.* Ai zheng = Aizheng = Chinese journal of cancer, 2007. **26**(1): p. 44-7.
133. Sarbia, M., et al., *Hypermethylation of tumor suppressor genes (p16(INK4A), p14(ARF) and APC) in adenocarcinomas of the upper gastrointestinal tract.* International Journal of Cancer, 2004. **111**(2): p. 224-228.
134. Mur, P., et al., *Impact on prognosis of the regional distribution of MGMT methylation with respect to the CpG island methylator phenotype and age in glioma patients.* J Neurooncol, 2015. **122**(3): p. 441-50.
135. Ahlquist, D.A., et al., *Next-Generation Stool DNA Test Accurately Detects Colorectal Cancer and Large Adenomas.* Gastroenterology, 2012. **142**(2): p. 248-256.
136. Wallner, M., et al., *Methylation of serum DNA is an independent prognostic marker in colorectal cancer.* Clinical Cancer Research, 2006. **12**(24): p. 7347-7352.
137. Deng, G.R., et al., *Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression.* Cancer Research, 1999. **59**(9): p. 2029-2033.
138. Widschwendter, M., et al., *Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen.* Cancer Research, 2004. **64**(11): p. 3807-3813.